

```
Db      1654 GGGCTGCTGCTCCAGGATTTTCAAGTCTCTAACAAGAGAGATGCAATTACCAAGT 1713
Qy      501 G1yG1uYrPjG1uValLeuG1yAqH1sG1ySerAn11eG1nVa1AaPjG1uValAqAq 520
Db      1714 GGTGAATGGAAAGTTCTTGCGCGGCGATGATCAATATCCAAAGTTGAAGTCAAGAG 1773
Qy      521 LeuValTyrPheG1uG1yThrTlyAaSPePLeuG1uH1sH1sLeuTyrValValSer 540
Db      1774 CTGGATATATTTTGAAGGACCAAGACCTCCCTTAGGACATCACTGTAGCTAGTCA 1833
Qy      541 TyrValAaSPePjG1uValValThrArgLeuThrAaPjG1yTyrSerH1sSerCyeCys 560
Db      1834 TACCTAAATCCTGGAGAGAGTACAAGGCTGACTGACCGTGGCTACTCACTTTGCTGC 1893
Qy      561 I1eSerG1nH1sCyAaSPePheH1eSerLysTyrSerAaG1nLysAaSPeH1sCyA 580
Db      1894 ATCAGTCAGCAGCTGACTCTTTATAGTAAGTACTAAACAGAAATCCACACACTGT 1953
Qy      581 ValSerLeuTyrLysLeuSerSerProG1uAaPjProThrCysLysThrLysG1uPhe 600
Db      1954 GTGTCCCTTACAGAGCTATCAAGTCTCAAGATGACCCCACTTGCAAAACAAAGGAATT 2013
Qy      601 TrpAlaThrI1eLeuAaSPeA1aG1yProLeuProAaPjTyrThrProProG1u1ePhe 620
Db      2014 TGGGCGACCATTTTGATTCAGCAGGTCCTCTCTGACTAATACCTCCCAAAATTTTC 2073
Qy      621 SerPheG1uSerThrThrGlyPheThrLeuTyrG1yWcLeuTyrLysProH1sAaPLeu 640
Db      2074 TCTTTGAAAGTACTACTGAGATTATCATTTGATGGAGATGCTCTCAACACCTCATGATCTA 2133
Qy      641 G1nProG1yLysLysTyrProThrValLeuPheH1eTyrG1yG1yProG1uValG1nLeu 660
Db      2134 CAGCGTGAAGAAATATCTCTACTGTGCTGTTCAATATAGGTGGTCTCTCAGGTGAGTTG 2193
Qy      661 ValAaSPeAaSPeH1ySg1yValLysTyrPheArgLeuAaThrLeuA1SerLeuG1y 680
Db      2194 GTGATATATCGGTTTAAAGAGAGTCAAGTATTTCCGCTTGAAATACCTTACCTCTAGCT 2253
Qy      681 TyrValValValVal11eAaSPeAaSPeG1ySerCyH1sAaG1yLeuLysPheG1uG1y 700
Db      2254 TATGTGCTGTGATGATGAGAAACAGGGGATCTGTCTCCAGGGCTTTAAATTTGAAAGC 2313
Qy      701 AlaPheLysTyrLysMetG1yG1n1eG1u1eAaPjAaPjG1nValG1yLysG1nTyr 720
Db      2314 GCCTTTAAATATTAATGGGTCMAANTAGAAATGACGATCGGTGGAAGGACTCCAAATAT 2373
Qy      721 LeuA1AaSPeArgTyrAaSPeH1eAaSPeLysAaPjValG1y1eH1sG1yTyrSerTyr 740
Db      2374 CTACCTTCTGATATGATTTTATTGACTTAGTGTGTGGGCACTCCAGCGTGGTCCAT 2433
Qy      741 G1yG1yTyrLeuSerLeuMetAlaLeuMetG1nArgSerAaP1ePheArgValA1a1e 760
Db      2434 GGAGAGATACCTCTCCCGATGGCATTAATGACAGAGTCAAGTATCTTCAGGGTGTGCTATT 2493
Qy      761 AlaG1yAlaProValThrLeuTriP1ePheTyrAaPThrG1yTyrThG1uArgTyrMet 780
Db      2494 GCTGGGCGCCCAAGTACCTCTGTGATCTTCTATGATACAGATACAGGAACTTATATG 2553
Qy      781 G1yH1sProAaPjAaSPeG1uG1nG1yTyrTyrLeuG1ySerValA1aMetG1nA1aG1u 800
Db      2554 GGTACACCTGACCAAGATGAACAGGGCTATTACTAGATCTGTGGCCATCAAGACGAA 2613
Qy      801 LysPheProSerG1uProAaPjArgLeuLeuLeuH1sG1yPheLeuAaSPjAaSPeVal 820
Db      2614 AAGTTCCTCTGAAACCAATGTTTACTGCTTACATGGTTCCTCGAAGAAATATC 2673
Qy      821 HisPheAaH1sThrSerI1eLeuLeuSerPheLeuValA1aG1yLysProTyrAaP 840
Db      2674 CATTTTGACATACACAGATATTACTGAGTTTATAGAGAGGGCTGAAACCATATGAT 2733
Qy      841 LeuG1n1eTyrProG1nG1uArgH1sSerI1eArgVal1ProG1uSerG1yG1uH1sTyr 860
```

```
Db      2734 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTCCTGMAATCGGAGACATTAT 2793
Qy      861 G1uLeuH1sLeuLeuH1sTyrTyrLeuG1nG1uAaSPeLysAaPjA1eA1a1eLys 880
Db      2794 GAATCGATCTTTTGCACTACCTTCAAGAAACCTTGATACGTAATGCTGCTTAAA 2853
Qy      881 Val1e 882
Db      2854 GTGATTA 2859

RESULT 6
AAH99934
ID AAH99934 standard; cDNA; 3143 BP.
AC AAH99934;
XX
XX
XX 12-APR-2002 (first entry)
DE cDNA encoding 21953 human prollyl oligopeptidase.
XX
XX 21953 prollyl oligopeptidase; human; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antichrithic; antiaesthetic; antiinflammatory;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
XX Homo sapiens.
XX 0
XX Key Location/Qualifiers
XX CDS 229..2877
XX FT /*tag= a
XX FT /product= "21953 prollyl oligopeptidase"
XX FT /note= "This region is specifically claimed in
XX FT Claim 2"
XX
XX MO200179473-A2.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001MO-US40483.
XX
XX 18-APR-2000; 2000US-197508P.
XX
XX (MILL-) MILENNIUM PHARM INC.
XX
XX Meyers RA, Williamson M;
XX
XX WPI; 2002-034353/04.
XX P-PSDB; AAG78415.
XX
XX New polypeptides 21953, member of human prollyl oligopeptidase family,
XX FT useful as diagnostic targets and therapeutic agents for controlling
XX PT cancer, lymphoma and leukemia
XX
XX
XX
XX Claim 7; Page 100-102; 121pp; English.
XX
XX This invention relates to an isolated 21953 human prollyl
XX CC oligopeptidase. Which is cytosolic, antidiabetic, antialthritic,
XX CC neuroprotective, antiepileptic, dermatological, antipsoriatic,
XX CC antiaesthetic, ophthalmological, antiinflammatory, nootropic,
XX CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
XX CC antiallergic, antiatherosclerotic, anorectic and
XX CC metabolic in its action. Uses include gene therapy, expression or
XX CC activity of 21953 protein modulator, it is useful for identifying a
XX CC compound which binds to it and can be used in preventing, treating
XX CC or degrading a cellular proliferative or differentiative disorder.
XX CC The 21953 molecules can act as novel diagnostic targets and therapeutic
XX CC agents for controlling disorders associated with the aberrant activity
XX CC or degradation of peptide hormones e.g., disorders associated with cell
XX CC differentiation and proliferation such as cancer, immune function,
XX CC reproductive, neurological and cardiovascular function. The 21953
XX CC molecules are thus useful for treating and preventing cellular
```

CC proliferative and differentiative disorders, haematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the cDNA
CC encoding sequence of 21953 human prolyl oligopeptidase.

XX Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	3143
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-070-464-1 (1-882) x AAH99934 (1-3143)

```
QY 1 MetAlaAlaIaMeGluThrGluGluGluValGluIlePheGluThrAlaAspCys 20
Db 229 ATGGCAGCAGCAATGGAAACAGAAACAGCTGGGTGTGAGATATTGAAACTGCCGACTGT 288
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
Db 289 GAGGAGAAATATTGAATACAGGATCGGCTTAATTGAGCCTTTTATTGTAAGCGGTAT 348
QY 41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 349 TCCCGAGTCAGCTTAAGAAAGCTCTTGCCATACCGAAATAATTCATGGTACATATAG 408
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 409 GCTAAGCACCACACTGATTTTCATGTTGTGAAGAGATATCCAGATGACCTCATTTCA 468
QY 81 AspArgLysTyrTyrLeuAlaMetSerGlyGluAsnArgLysGluAsnThrLeuPheTyrSer 100
Db 469 GACAGAAATCATTAACCTTGGCATGTCTGGTGAGAAACAGAAATAACACTGTTTATTTCT 528
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db 529 GAAATTCCTCAAAACTATATCAATAGACAGACGCTTAATAGCTCTTGGAAGCCCTTTTGG 588
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db 589 GATCTTTTTCAGGCAACACTGGACTATGGAATGATTCGAGAGAAAGAACTATTAGA 648
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db 649 GAAAGAAACGCAATGGAAACAGTCGGAAATGCTTCTTACGATTAATCACCAAGGAAGTGA 708
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspArgLysProGlnGly 180
Db 709 ACATTTCTGTTTCAAGCCGGAGTGGAAATTAATCACTAAAGATGGAAGGCGCCCAAGGA 768
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 769 TTTACGCAACACTTTTAAGGCCCAATCTAGTGAATCTAGTTGTCCCAACATACGGAATG 828
QY 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db 829 GATCCAAATTAATGCTCTGCTGATCCAGACTCGATTCCTTTTATATACCAACAGATATT 888
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db 889 TGGATATCTAACATCTGTAACCAAGAGAAAGAGACTCACTTAATGCGCAATGAGCTA 948
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db 949 GCCAACATGGAAAGATGCGCAGATCAGTGAAGTGGCTACTTTGTTTCCAAAGAAAGAA 1008
QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280
|||||
```

```
Db 1009 TTGTATGATTAATTCGTGCTATTGGTGGTGTCCAAAGCTGAAACAACTCCAGTGTGCT 1068
QY 261 LysIleLeuArgIleLeuTyrGluAsnAspGluSerGlyValGluIleIleHisVal 300
Db 1069 AAAATTCCTTGAATTCATATGAAAGAAATGATAACTGGGTGGAAATTAATTCATGTT 1128
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 1129 ACATCCCTATGTTGGAAACAGAGAGGCAATTCATTCCTTATCCTAACAGATACA 1188
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db 1189 GCAATTCCTAAAGCACTTTTAAGATGTCAGAAATATGATGATGCTGGAAGGAAGATC 1248
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1249 ATAGATGTCAATAGTAAGAACTAATTAACCTTTTATGAGATTCATTAATGGAAGGATGAA 1308
QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db 1309 TATATTCCAGAGCTGATGAGACTCCGAGGAGAAATATGCTTGTCATCTCATAGAT 1368
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db 1369 CGCTCCAGACTCGCTGCAAGATAGTGTGATCTCACTGAATTAATTATCCAGTAGAA 1428
QY 401 AspAspValMetGluValArgGluIleGluSerValProAspSerValThrProLeu 420
Db 1429 GATGATGTATGAAAGGAGAGACATTCATGAGCAAGTGGCTGTGTCGAGCCCACTA 1488
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db 1489 ATTAATCTTAAGAAACAAACAGACATGTGATTAATTCATGACATCTTTCATGTTTTT 1548
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db 1549 CCCCAAGTCACAGAAAGAAATGAGTTATTTTCTCTGCTGAATCCAAACACGATTC 1608
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerTyrTyrLysArgSerSerGly 480
Db 1609 CGTCATTTAATCAAAATTAATCAATCTATTTTAAGGAAAGCAAAATTAACATCCAGTGT 1668
QY 484 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db 1669 GGGCTGCTCTCCAAAGTATTTCCAGTGTCTCATCAAAAGAGATAGCAATTAACAGT 1728
QY 501 GlyLeuTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db 1729 GTGGAATGGGAAGTTCTTGCCCGCATGATCTAATATCCAAATGTAAGTCAGAAAG 1788
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisValLeuValSer 540
Db 1789 CTGTATTAATTTTAAGGACAAAGCACTCCCTTTAGAGATCACTCGTACGATGACGT 1848
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrThrHisSerCysCys 560
Db 1849 TACGTAAATCTTGAAGAGTGAACAGGCTGACGACCGTGCATCTACATTCCTGCTGC 1908
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db 1909 ATCAGTCAGACTGTGACTTTTATTAATAGTAATTAATGTAACAGAAATCCACTGT 1968
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db 1969 GTGTCCCTTTAAGATATCAACGTCTGGAAGATGACCACCTTGCAAAACAAAGGAATTT 2028
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db 2029 TGGGCAACATTTTGAATTCAGAGAGTCTCTTCTGACATATCTCTCCAGAAATTTTC 2088
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db 2089 TCTTTTGAAGATCTACTGATTAATGATTAATGAGATGCTTACAAAGCTCATGATCTA 2148
|||||
```



```

Y      641 GlnProGlyLysValSerTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
      2149 CACCTGGAAAGAAATATCTTACTGCTGCTGTCATATATGTCCTCCCTCAGGTCAGTTG 2208
Y      661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnPhrLeuAlaSerLeuGly 680
      2209 GTGAATTAATCGTTTAAAGAGTCAGATATTTCCGCTTGAAATACCTTACGCTCTAGGT 2268
Y      681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGly 700
      2269 TATGTGGTTGATGATAGACACACAGGGATCTCTGACCGAGGGCTTAATTGAAAGGC 2328
Y      701 AlaPheLysTyrLysMetGlyGlnIleGlnIleAspAsnGlnValGlnGlyLeuGlnTyr 720
      2329 GCCTTTAAATTAATAATGGTCAATAAGAAATGACATGACATGAGTGAAGGACCTCAATAT 2388
Y      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
      2389 CTAGCTTCGATATGATTTCTATGATTAATGATGATGATGATGATGATGATGATGATGATG 2448
Y      741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
      2449 GGAGGATACCTCTCCCTGATGAGCATTAATGACAGAGTCAGATATCTTCAGGGTCTATTT 2508
Y      761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlnArgTyrMet 780
      2509 GCTGGGGCCCCAGCTCCTGTCGATCTTCTATGATGATGATGATGATGATGATGATGATG 2568
Y      781 GlyHisProAspGlnAsnGlnGlnGlyTyrTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
      2569 GGTTCACCTCGACCGAATGATGACAGGCTATTTACTTACTTACTTACTTACTTACTTACTT 2628
Y      801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
      2629 AAGTTCCTCCCTGAAACCAATTCGTTTACTCTCTTACTGATGATGATGATGATGATGATG 2688
Y      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaArgAlaGlyLysProTyrAsp 840
      2689 CATTTTCACATACCACTATATTTACTGATGATGATGATGATGATGATGATGATGATGATG 2748
Y      841 LeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGluSerGlyGlnHisTyr 860
      2749 TTACAGATCTATCTCTCGAGAGACACAGCATTAAGATTCCTGAAATGGGAGAACATTAT 2808
Y      861 GlnLeuHisLeuLeuHisTyrLeuGlnGlnGlnAsnLeuGlySerArgIleAlaAlaLeuLys 880
      2809 GAACCTGCATCTTTTGCACTACCTTCAAGAAAACCTTGATCAGATGCTGCTCTAATA 2868
Y      881 ValIle 882
      2869 GTGATTA 2874
Db

```

RESULT 7

AAH99935 standard; cDNA; 2643 BP.

AAH99935:

12-APR-2002 (first entry)

Coding sequence of 21953 human prolyl oligopeptidase.

21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
 cancer; cardiovascular disease; autoimmune disease; atopic allergy;
 neuronal disorder; vascular disorder; prostate disorder; cytostatic;
 antidiabetic; anticholinergic; antiaesthetic; antiinflammatory;
 diabetes mellitus; arthritis; multiple sclerosis; asthma;
 Grave's disease; neuronal disorder; demyelinating disease; ss.

XX Homo sapiens.
 OS
 XX
 PN WO200179473-A2.

```

XX      25-OCT-2001.
PD      11-APR-2001; 2001WO-US40483.
XX      18-APR-2000; 2000US-197508P.
XX      (MILL-) MILLENNIUM PHARM INC.
XX      Meyers RA, Williamson M;
XX      WPI: 2002-034353/04.
XX      P-PSDB; AAG78415.
XX      New polypeptides 21953, member of human prolyl oligopeptidase family,
XX      useful as diagnostic targets and therapeutic agents for controlling
XX      cancer, lymphoma and leukemia
XX      Claim 2: Page-; 121pp; English.
XX      This invention relates to an isolated 21953 human prolyl
XX      oligopeptidase, which is cytostatic, antidiabetic, antiarthritic,
XX      neuroprotective, antihypertensive, dermatological, antipsoriatic,
XX      antiaesthetic, ophthalmological, antiinflammatory, nootropic,
XX      antiparkinsonian, anticonvulsant, gynaecological, vasorelaxant,
XX      antidiabetic, cardiac, antithrombotic, anorectic and
XX      antitumor, in its action. Uses include gene therapy, expression or
XX      metabolic in its action. Uses include gene therapy, expression or
XX      activity of 21953 protein modulator, it is useful for identifying a
XX      compound which binds to it and can be used in preventing, treating
XX      or detecting a cellular proliferative or differentiative disorder.
XX      The 21953 molecules can act as novel diagnostic targets and therapeutic
XX      agents for controlling disorders associated with the aberrant activity
XX      or degradation of peptide hormones e.g., disorders associated with cell
XX      differentiation and proliferation such as cancer, immune function,
XX      reproductive, neurological and cardiovascular function. The 21953
XX      molecules are thus useful for treating and preventing cellular
XX      proliferative and differentiative disorders, haematopoietic neoplastic
XX      disorders, immune disorders such as autoimmune diseases, diabetes
XX      mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
XX      neuronal disorders, demyelinating diseases, vascular disorders and
XX      metabolic or pain disorders. This sequence represents the cDNA
XX      encoding sequence of 21953 human prolyl oligopeptidase. This
XX      sequence represents the coding sequence of 21953, being the sequence
XX      in between the start and the stop codon of the sequence represented in
XX      AAH99934.
XX      Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;
XX      Alignment Scores:
XX      Pred. No.: 0 Length: 2643
XX      Score: 4695.00 Matches: 881
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 99.89% Indels: 0
XX      DB: 24 Gaps: 0
XX      US-10-070-464-1 (1-882) x AAH99935 (1-2643)
Y      2 AlaAlaIleMetGluThrGlnGlnLeuGlyValGlnIlePheGluThrAlaAspCysGlu 21
      1 GCGACGACGATGGAAGAAACAGACGCTGGGTGATGATTAATGAAACATGCGACATGAG 60
Y      22 GlnAsnIleGluSerGlnAspArgProLysLeuGlnPhePheTyrValGlnArgTyrSer 41
      61 GAGAAATATTAATGATCAACAGATGCGCTTAATGAGCCCTTTTATGTTGAGCGGATTC 120
Y      42 TTPSerGlnLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMetAla 61
      121 TGGAGTCACCTTAAGAAAGCTGCTTCCGATACACAGAAATATCATGCTACATGATGCT 180
Y      62 LysAlaProHisAspPheMetPheValLysArgAsnArgProAspGlyProHisSerAsp 81
      181 AAGGACACCATATTTCTATGTTTGAAGAGGATGATGATGATGATGATGATGATGATG 240
Db

```

OY	82	ArgIIeTYrTYrLeuAlaMeSeRGylVGIuenaRGluAsnThrLeuPheTYSeRGlu	101
Db	241	AGAACTCTATACCTTGGCCATGTCTGGTGAAACAGAGAAAATACACTGTTTATTCCTGAA	300
OY	102	IIeProLYsThrIIeAsnArgAlaAlaValIIeMeLSeTYrPlySProLeuAsp	121
Db	301	ATTCGCAAAACATATCATATGACACAGACTTAAAGCTCTTTGGAAAGCTTTTGGAT	360
OY	122	LeuPheGlnAlaThrLeuAspTYrGlyMeTYrSeRArgGluGluLeuLeuArgGlu	141
Db	361	CTTTTTCAGGCAACACTGGACATATGAAAGTATTCCTCGAAGAAAGAACTATTAAAGACA	420
OY	142	ArgLYsArgGIIeGlyThrValGlyIIeAlaSeTYrAspTYrHisGIIeGlySeRGlyThr	161
Db	421	AGAAACCGATTGGAAACAGTCGGAAATGCTTCTTACGATTATCACCAAGAAAGTGAAACA	480
OY	162	PheLeuPheGlnAlaGlySeRGlyIIeTYrHisValLYsAspGlyGlyProGIIeGlyPhe	181
Db	481	TTTCTGTTTCAAGCGCGTAGTGGAATTTATTCAGTAAAGATGGAAGGCGCCACAAGGATTT	540
OY	182	ThrGIIeGlnProLeuAspProAsnLeuValGluThrSeCYsProAsnIIeArgMeLasp	201
Db	541	ACGCACAACCTTTTAAAGGCCCAATCTAGTGAAGACTAGTTGTCACCAACTAGCAGATGAT	600
OY	202	ProLYsLeuCYsProAlaAspProAspTYrPIIeAlaPheIIeHisSeRAsnAspIIeTYrP	221
Db	601	CCAAATATTATGGCCCGCTGATCCAGACTGGAATTCCTTTTATATCATAGCAACATATTGG	660
OY	222	IIeSeRAsnIIeValIIhThrArgGluGluArgArgLeuThrTYrValHisAsnGluLeuAla	241
Db	661	ATATTATACATCTGTAAACAGAGAAAGAAAGAAACCTACTTATGTGCACATATAGCTAGCC	720
OY	242	AsnMeTGluGluAspAlaArgSeRArgIAGlyValAlaThrPheValLeuGIIeGluPhe	261
Db	721	AACATGAAAGANGATGCCAGATCAGCTGGAGTCGTAACCTTGTGTCTCCAAAGAAATTT	780
OY	262	AspArgTYrSeGlyTYrTYrTYrPProCYsProLYsAlaGluThrThrProSeRGlyLYs	281
Db	781	GATAGATATTCTGGCTATTGGTGGTGCCAAAGCTGAACAACTCCGAGTGCGTAA	840
OY	282	IIeLeuArgIIeLeuTYrGluGluAsnAspGlySeRGlyValGluIIeIIeHisValThr	301
Db	841	ATTCTTAAGATTCTATATGAAAGAAATGATGTAATCGAGTGAAATTAATCATATTTCACA	900
OY	302	SeRPrometLeuGluThrArgArgAlaAspSeRPhaArgTYrProLYsThrGIIhThrAla	321
Db	901	TCCCTTATGTTGGAAACAAAGAGGGCAGATTATTCGGTATCTTAAACAGGTRACAGA	960
OY	322	AsnProLYsValThrPheLYsMeSeRGluIIeMeTIIeAspAlaGluGlyArgIIeIIe	341
Db	961	AATCTTAAGCTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGAAAGATCATTA	1020
OY	342	AspValIIeAspLYsGluLeuIIeGlnProPheGIIeLeuPheGluGlyValGluTYr	361
Db	1021	GATGTCATATGATMAAGAACTAATTCAACTTTTGATTCATATTTGAAGAGGTTGAATAT	1080
OY	362	IIeAlaArgAlaGlyTYrThrProGIIeGlyLYsTYrAlaTYrSeRIIeLeuLeuAspArg	381
Db	1081	ATTGCGCAAGCTGATGATGACTCTGAGGGAATAATGCTGGTCCATCTACTATGATGCC	1140
OY	382	SeRGlnThrArgLeuGlnIIeValLeuIIeSeRProGluLeuPheIIeProValGluAsp	401
Db	1141	TCCCAAGCTCCCGTCAGATATGATGTGATCTCACTCGAATTAATTTATCCCACTAAGAAAT	1200
OY	402	AspValMeTGluArgGlnArgLeuIIeGlySeRValProAspSeRValThrProLeuIIe	421
Db	1201	GATGTTATGGAAGCAAGCACAAGACTATTGATGATGAGCGCTGATTCGTGACGCACTAATT	1260
OY	422	IIeTYrGluGluThrThrAspIIeTYrPIIeAsnIIeHisAspIIePheHisValPhePro	441
Db	1261	ATCTATAGAGAAACAACAGACTCTGGATTAATATATCATATGATCTTTCAATGTTTCTCC	1320

QY	442	GINSEHRIAGLUGLUGLIUEGLUEHIEPEALASERGLUYELVETHRTGLYHEARG	461
DB	1321	CAAAAGTCAGAAAGAGAAATTGAGTTTATTTTGGCTCGAATGCAAAACAGATTTCCT	1380
QY	462	HI ⁶ LEU ¹ TYR ¹ LYS ¹ ET ¹ HSR ¹ ILEU ¹ YSGLUSELTYR ¹ TYR ¹ LYS ¹ ASR ¹ SER ¹ GLY ¹	481
DB	1381	CA ¹ TTTAT ¹ TACAA ¹ ATT ¹ ACAT ¹ CTAT ¹ TTTAA ¹ AGAA ¹ GCMA ¹ TAT ¹ AA ¹ GCAT ¹ CCAGTGGG	1444
QY	482	LEUPROALAPROSERASPHELYSCYSPROILETVSGIUGLIUEALILE ¹ THSRGLY	501
DB	1441	CGGCTCGCCCAAGGATTTTCAAGTCTCTATCAAAAGAGATGCAATTAACGTGT	1500
QY	502	GIUT ¹ RGIV ¹ AL ¹ LEU ¹ GLY ¹ ARG ¹ HIS ¹ GLY ¹ SE ¹ ASN ¹ ILE ¹ GIN ¹ VAL ¹ ASP ¹ GLVAL ¹ ARG ¹ LEU	521
DB	1501	GAATGGGAAGTCTTGCGCGCATGATGATTAATATCCAAAGTTGAAGAGTCCAGAGCGCTG	1566
QY	522	VAL ¹ TYR ¹ PHE ¹ GLUG ¹ LY ¹ THR ¹ LYS ¹ ASP ¹ SE ¹ PRO ¹ LEU ¹ HI ¹ SHILE ¹ TYR ¹ VAL ¹ SER ¹ TYR	541
DB	1561	GTA ¹ TAT ¹ TTTGAAGGACCAAAACATCCCTTTAAGGATCACTGTACGTGATCGTTAC	1620
QY	542	VAL ¹ ASP ¹ PRO ¹ GLY ¹ VAL ¹ THR ¹ ARG ¹ LEU ¹ HR ¹ AP ¹ ARG ¹ LY ¹ TYR ¹ SER ¹ HIS ¹ GLY ¹ LE	561
DB	1621	GTAAATCCCGAGAGAGTGCACAAAGCTGACTGACCGTGCATCTCATTTCTGCTGCATC	1688
QY	562	SERGINHIS ¹ CYASAPSPHEHIE ¹ SER ¹ LYS ¹ SE ¹ ASN ¹ GLY ¹ ASP ¹ PROHIS ¹ CYS ¹ VAL	581
DB	1681	AGTCAGCACTGATCTTCTTTATAGTATAGTAA ¹ CCAGAAATCCACACTGTGTG	1740
QY	582	SERLEU ¹ TYR ¹ LYS ¹ LEU ¹ SER ¹ PRO ¹ GLY ¹ ASP ¹ PRO ¹ THR ¹ CY ¹ ELV ¹ ETH ¹ LYSGIUEHTRP	601
DB	1741	TCCCTTTACAACTATCAAGCTCTGMAATGACCCCACTCTCCAAACAAAGAAATTTTGG	1800
QY	602	ALATHRIILEUASPSE ¹ ALGLY ¹ PRO ¹ LEUPRO ¹ ASP ¹ TYR ¹ THR ¹ PRO ¹ GLY ¹ LEPHE ¹ SER	621
DB	1801	GCCACCATTTGGATTTCCAGAGTCTCTTCTCTACTATCTACTCTCCGAAATTTTCTCT	1866
QY	622	PHEGLUSE ¹ THR ¹ THGLY ¹ PHE ¹ THR ¹ LEU ¹ TYR ¹ GLY ¹ ME ¹ LEU ¹ TYR ¹ LYS ¹ PROHIS ¹ ASP ¹ LEU ¹ GIN	641
DB	1861	TTTGAAGAAGTACTGATGATTAATCATGTATGGAGAGCTCTAACAGCTCATGATCTTACAG	1920
QY	642	PROGLY ¹ LYS ¹ LYS ¹ TYR ¹ PRO ¹ THR ¹ VAL ¹ LEU ¹ PHE ¹ ILE ¹ TYR ¹ GLY ¹ GLY ¹ PRO ¹ GIN ¹ VAL ¹ GLY ¹ LEU ¹ VAL	661
DB	1921	CCGTGAAAGAAATATCCATCTGTCTGTCTATATGTGTGGTCCACAGTGCAGTTGTGTG	1980
QY	662	ASPANASPHELYSGI ¹ VAL ¹ LYS ¹ TYR ¹ PHE ¹ ARG ¹ LEU ¹ ASN ¹ THR ¹ LEU ¹ ALASER ¹ LEU ¹ GLY ¹ TYR	681
DB	1981	AATTAATCGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCCTTACCTCTCTAGGTAT	2040
QY	682	VAL ¹ VAL ¹ VAL ¹ ILE ¹ ASP ¹ ASN ¹ ARG ¹ LYS ¹ CYSHIS ¹ ARG ¹ LY ¹ LEU ¹ LYS ¹ PHE ¹ GLY ¹ VAL	701
DB	2041	GTGGTGTAGTATGATGACACAGGGAGCTCTGTCAACGAGGCTTAAATTTGAAGCCCC	2100
QY	702	PHELYSTYR ¹ LYS ¹ ME ¹ GLY ¹ GIN ¹ ILE ¹ GIN ¹ LEASP ¹ ASP ¹ GIN ¹ VAL ¹ GLY ¹ LEU ¹ GIN ¹ TYR ¹ LEU	721
DB	2101	TTTAATATTA ¹ AAATGGGTCAATATGAAATTAACGATCAGGTGAAAGCACTCCATATCTA	2160
QY	722	ALASER ¹ ARG ¹ TYR ¹ ASP ¹ PHE ¹ ILE ¹ ASP ¹ LEU ¹ SP ¹ ARG ¹ VAL ¹ GLY ¹ ILE ¹ HIS ¹ GLY ¹ TRN ¹ SER ¹ TYR ¹ GLY	741
DB	2161	GCTTCTCATATGATTTATGACTTATGATGCTGTGGGCATCCACGCGTGTCTTATGGA	2220
QY	742	GLY ¹ TYR ¹ LEU ¹ SER ¹ LEU ¹ ME ¹ AL ¹ LEU ¹ ME ¹ GIN ¹ RG ¹ SER ¹ AP ¹ ILE ¹ PHE ¹ ARG ¹ VAL ¹ ILE ¹ AL	761
DB	2221	GGATACCTCTCCCTGATGGCATTAATGACAGGTGACGATATCTTCAGGGTTGCTATTTGCT	2280
QY	762	GLY ¹ AL ¹ AP ¹ PRO ¹ VAL ¹ THR ¹ LEU ¹ TRP ¹ LEPHE ¹ TYR ¹ ASP ¹ THR ¹ GLY ¹ TYR ¹ TH ¹ GIN ¹ ARG ¹ TYR ¹ ME ¹ GLY	781
DB	2281	GGGGCCCAAGTACACTGTGATCTTCTATGATCAGAGTACACGGAACGTTATATGGGT	2340
QY	782	HIS ¹ PRO ¹ ASP ¹ GIN ¹ ASN ¹ GLY ¹ GIN ¹ GLY ¹ TYR ¹ TYR ¹ LEU ¹ GLY ¹ SER ¹ VAL ¹ AL ¹ ME ¹ GIN ¹ AL ¹ GLY ¹ YS	801
DB	2341	CACCTTGACCAAGATGAACAGGCGCTATTACTTACGATCTGTGGCATGACCAAGCAAGAAAG	2400

DB 2401 TTCCCTCTGGAACCAATCGTTACTGCTTACATGGTTTCCGGAAGAAATGTCAT 2460
QY 822 PheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGAlaGlyLeuProTyrAspLeu 841
DB 2461 TTTCACATACACAGATATTTACTGAGTTTAAAGAGGCGTGAACCATATCATTTTA 2520
QY 842 GlnIleTyrProGlnGluAlaGHisSerIleArgValProGluSerGlyGlnHisTyrGln 861
DB 2521 CAGATCTATCTCTCAGAGAGACACAGCATTAAGAGTCTCGAATCGGAGAACATTATGAA 2580
QY 862 LeuHisIleLeuHisIleTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuIysVal 881
DB 2581 CTGCATCTTTTGCACTACCTTCAGAAAACCTTGATCAGATATGCTGCTTAAAGTG 2640
QY 882 Ile 882
DB 2641 ATA 2643
RESULT 8
ABK83327
ID ABK83327 standard; cDNA; 4829 BP.
XX
AC ABK83327;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-1 splice variant #3.
XX
KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyslexia; reproductive disorder; inflammatory disorder;
KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN MO2002J1134-A2.
XX
PD 18-APR-2002.
XX
PE 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
XX MPI; 2002-444178/47.
DR P-PSDB; ABG61596.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS -
XX
XX Disclosure; Page 65-66; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
XX proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
XX proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
XX and nucleic acids encoding them are useful for treating infections
XX such as fungal, bacterial, protozoan and viral infections, particularly
XX infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
XX pain, diabetes, precocious puberty, infertility, obesity, anorexia,
XX bulimia, Parkinson's disease, acute heart failure, hypotension,
XX hypertension, urinary retention, osteoporosis, angina pectoris,
XX stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
XX psychotic and neurological disorders (e.g. anxiety, dementia, or
XX schizophrenia), and dyslexias. These may also be used in discovering

CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ Sequence 4829 BP; 1466 A; 886 C; 1017 G; 1460 T; 0 other;
Alignment Scores:
Pred. No.: 0 Length: 4829
Score: 4680.00 Matches: 882
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.57% Indels: 2
DB: 0 Gaps: 0
US-10-070-464-1 (1-882) x ABK83327 (1-4829)
QY 1 MetaAlaAlaIleMetGluThrGluGlnIleuGlyValGluIlePheGluThrAlaAspCys 20
DB 214 ATGGCAGCAGCAGCATGGAAGAGAAAGAGTGGGTGTGAGATATTGGAACGCGCAGCTGT 273
QY 21 GluGluAsnIleGluSerGlnAspArgProIysLeuGluProPheTyrValGluArgTyr 40
DB 274 GAGGAGATATTGGAATCAGAGATCGGCTAAATTGAGCCTTTTATGTTGAGCGGTAT 333
QY 41 SerTrpSerGlnLeuIysIleuValAspThrArgIleTyrHisGlyTyrMetMet 60
DB 334 TCCTGAGTCAGCTTAAAGAGCTGCTGCGCATACCAAAAATATCATGCGCTACATGATG 393
QY 61 AlaIysAlaProHisAspPheMetPheValIysArgAsnAspProAspGlyProHisSer 80
DB 394 CTTAAGCAGCAGCAGCATGATTTTCATGTTTGTGAAGAGATGATCCAGATGAGACTATTCA 453
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 454 GACAGAAATCTATTACCTTGCATGCTGTGTGAGAACAGAGAAAATACACTGTTTATTCT 513
QY 101 GluIleProIysThrIleAsnArgAlaIleValIleuMetLeuSerTrpIysProIleu 120
DB 514 GAAATTCCTCAAAACATATGATGAGCAGCAGATCTTAATGCTCTTGGAGGCTCTTTTG 573
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluIleuLeuArg 140
DB 574 GATCTTTTTCAGGCAAGCTGAGACTGAGATGATTTCTCGAGAAAGAACTATTAAAG 633
QY 141 GluArgIysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnIysGly 160
DB 634 GAAAGAAAACCATTTGGAACAGTCGGAATTCCTTTCGATTTATCACCAAGGAAGTGA 693
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValIysAspGlyGlyProGlnGly 180
DB 694 ACATTTCTGTTTCAAGCCGGTAGTGAAATTATATCGTAAAGATGAGGCGCACAGGA 753
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 754 TTTACGCAACAACCTTTAAGGCCCAATCTATGTGGAATAGTTGCCAACATAGGATG 813
QY 201 AspProIysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
DB 814 GATCCAAAATTATATGCGCTGATCCAGATCCAGATTCGTTTATATCATAGCAACATAT 873
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgIleuThrTyrValHisAsnGluLeu 240
DB 874 TGGATATCTTAACATCGTAACAGAGAAAGAGAGACCTCATATGTCACACATAGCTA 933
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGln 260
DB 934 GCCAACATGGAAGAGAGATGCCAGATCGTAGTGGTACTTTGTTCTCCAGAAAGAA 993
QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProIysAlaGluThrThrProSerGlyGly 280
DB 994 TTTGATATATTTCTGCGTATTTGGTGTGTCCAAAAGCTGAAACCACTCCACAGTGCTGT 1053
QY 281 LysIleLeuArgIleuTyrGluGluAsnAspGluSerGlyValGluIleHisVal 300

```

Db      1054 AAAATCTTAGAATTCTATATGAGAAATGATGATCTGAGTGGAAATTATTCATGTT 1113
Qy      301  ThSerProMetLeuGluThrArgArgAlaAspSerPheArgTyPProLeuThGlyThr 320
Db      1114 ACATCCCTATGTTGGAAACAGAGGAGATTCATTCGGTTATCTTAAACAGGTACA 1173
Qy      321  AlaAspProLyValThrPheLyMetSerGluIleMetIleAspAlaGluIleArgIle 340
Db      1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATTAATGATTGATGCGAAGAAAGATC 1233
Qy      341  IleAspValIleAspLyGluLeuIleGlnProPheGluIleLeuPheGluIleValGlu 360
Db      1234 ATAAATGTCATAGATTAAGAACTAATTCACCTTTTGATTCATTTTGAAAGGAGTTGAA 1293
Qy      361  TyrIleAlaArgAlaGluTyrThrProGluIleTyPProIleAspSerIleLeuLeuAsp 380
Db      1294 TATATTCGCAAGCTCGATGAGTCTCTGAGGAAATATGCTTGCTTCATCTTACTAAT 1353
Qy      381  ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db      1354 CGCTCCGAGACTCGCCTACAGATAGTGTGATCTCACCTGAATTATTATCCAGTAGAA 1413
Qy      401  AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db      1414 GATGATGTTATGGAAGGACAGAGACTCATGTGACAGTCTGATTTCTGTGACGCACTA 1473
Qy      421  IleIleTyrGluGluThrThrAspIleTyrIleAspIleIleAspIlePheIleValPhe 440
Db      1474 ATTAATCTATGAGAAACAAACAGACATCTGATTAATATCCATGACATCTTTCATGTTT 1533
Qy      441  ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysValThrGlyPhe 460
Db      1534 CCCCAAGTCAAGAAAGAAATGAGTTATTTTGGCTCTGAAATGAAAACAGTTTC 1593
Qy      461  ArgHisLeuTyrLySileThrSerIleLeuLySerLySileTyrLyAspSerSerGly 480
Db      1594 CGTCATTTATACAAATTTACATCTATTTTAAAGAAACAAATTAACGATCCACAGTGT 1653
Qy      481  GlyLeuProAlaProSerAspPheLySAspProIleLySglnGluIleAlaIleThrSer 500
Db      1654 GGGCTGCTGCTCCAGATGATTTCAAGTCTCTTCAAGAGAGATGCAATTAACAGT 1713
Qy      501  GlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db      1714 GGTGAATGGGAAGTCTTGGCGGCAATGATATATCCAAAGTTGAATGAAAGTCAAGAG 1773
Qy      521  LeuValIlyrPheGluGlyThrLyAspSerProLeuGluHisIleSleuTyrValIleSer 540
Db      1774 CTGGTATATTTTGAAGGACCAAAAGACTCCCTTTAGGACATCACTGTAGTACAGTCA 1833
Qy      541  TyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db      1834 TACGTAATCTCGAGAGGTACAGAGGCTACCTACCGGTGCTACATCTTGTGCTGC 1893
Qy      561  IleSerGlnHisCysAspPhePheIleSerLySileTyrSerAsnGlnLyAsnProHisCys 580
Db      1894 ATCAGTCAGACCTGTGACTTTTATAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1953
Qy      581  ValSerLeuTyrLySleuSerSerProGluAspAspProThrCysValThrLySglnPhe 600
Db      1954 GTGTCCCTTTACAGATCATCAAGTCTGAGATGACCAACCTTCAAAACAAAGAAATTT 2013
Qy      601  TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db      2014 TGGGCGCCCATTTTGGATTCAGCGAGGCTCTCTCTGATCTATCTCTCCCGAAATTTTC 2073
Qy      621  SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLySProHisAspLeu 640
Db      2074 TCTTTTGAAGTACTACTGATTTACATTTGATGAGATGCTTCAAGGCTCAGATCTA 2133
Qy      641  GlnProGlyLySlySlyTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGln 660
Db      2134 CAGCCTGGAAAGAAATCTCTACTGTCTGTTTATATATGATGATGCTCTCAGGTCAGT 2193

```

```

Qy      660  euValAsnAsnArgPheLySglnValLySlyrPheArgLeuAsnThrLeuAlaSerLeuG 680
Db      2194 TGGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTAGCTTCTAG 2253
Qy      680  IlyTyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLySProPheGluG 700
Db      2254 GTTATGCTGTTGTAGTATACAAACAGGGATCTCTGACCGAGGGCTTAAATTTGAAAG 2313
Qy      700  IlyAlaPheLySlyrLySMetGlyGlnIleGluIleAspAspGlnValGluIleLeuGlnT 720
Db      2314 GCGGCTTTAAATTAATAATGGGTCAATATGAATTTGAAGATGATGAGTGAAGACCTCAAT 2373
Qy      720  TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTPSERT 740
Db      2374 ATCTAGCTTCTCGATATATTCATTTGATGATGATGATGATGATGATGATGATGATGAT 2433
Qy      740  TyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760
Db      2434 ATGAGAGTACCTCTCCCTGATGGCATTAATGCAAGAGTCAAGATATCTTCAAGGCTGCTA 2493
Qy      760  leAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrM 780
Db      2494 TTGCTGGGGCCCGAGTCACTCTGTGATCTTCTATGATACAGATACAGGAACGTTATA 2553
Qy      780  etGlyHisAspAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaG 800
Db      2554 TGGGTCACTCCGACAGATGAACAAGGCTATTAATCTTAAGATCTGTGGCCATGCAAGAG 2613
Qy      800  IlyLySProSerGlyProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnV 820
Db      2614 AAAAGTTCCTCTGACAAACAAATGCTTACGCTCTTACATGCTTCTGATGAGAAATG 2673
Qy      820  AlHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLySProTyrA 840
Db      2674 TCCATTTTGCATATACCAAGTATATTAATGATGATTTTGTAGAGGCTGAAAGCCATATG 2733
Qy      840  sPLeuGlnIleTyrProGlnIlyrArgHisSerIleArgValProGluSerGlyGluHisT 860
Db      2734 ATTAACAGATATATCTCGAGAGAGACACAGCATTAAGATTCCTGAATCGGAGAAACAT 2793
Qy      860  yrcGluLeuHisSleuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeu 880
Db      2794 ATGAACGCAATCTTTTGCATCTTCAAGAAACCTTGGATCAAGTATGCTGCTCTAA 2853
Qy      880  ySValIle 882
Db      2854 AAGTGATA 2861

```

RESULT 9
 ID ABRK83332 standard; cDNA; 4685 BP.
 AC ABRK83332;
 DT 12-AUG-2002 (first entry)
 XX cDNA encoding human DPP-1 splice variant #8.
 DE
 OS Homo sapiens.
 XX
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX

PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR MPI: 2002-444178/47.
XX P-PSDB; AB661601.
PT
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS Disclosure; Page 75-76; 113pp; English.

XX The present invention relates to the isolation of novel human serine
XX proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
XX proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
XX and nucleic acids encoding them are useful for treating infections
XX such as fungal, bacterial, protozoan and viral infections, particularly
XX infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
XX pain, diabetes, precocious puberty, infertility, obesity, anorexia,
XX biliary, Parkinson's disease, acute heart failure, hypotension,
XX hypertension, urinary retention, osteoporosis, angina pectoris,
XX stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
XX psychotropic and neurological disorders (e.g. anxiety, dementia, or
XX schizophrenia), and dyskinesias. These may also be used in discovering
XX and therapeutic agents for the treatment of reproductive, inflammatory and
XX metabolic disorders. ABK8332-ABK83343 encode human DPRP proteins.

XX Sequence 4685 BP; 1430 A; 853 C; 991 G; 1411 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4685
Score: 4385.50 Matches: 834
Percent Similarity: 94.56% Conservative: 0
Best Local Similarity: 94.56% Mismatches: 1
Query Match: 93.31% Indels: 48
DB: 24 Gaps: 1

US-10-070-464-1 (1-882) x ABK83332 (1-4685)

OY 1 MetAlaAlaAlaMetGluThrGluGluGluValGluIlePheGluThrAlaAspCys 20
DB 214 ATGCAGACGACGAAATGGAACAGACAGCTGGGTGTGATGATTTGAAATCTCGGACTGT 273
OY 21 GluGluAsnIleGluSerGluAsnAspProGluGluPheGluValGluArgTyr 40
DB 274 GAGGAGAAATGTAATCAGACAGATCGGCTTAAATGAGCCTTTTATGTTGAGCGGTAT 333
OY 41 SerTSPSeGluLeuGluLeuValAspThrArgLysTyrIleGluTyrMetMet 60
DB 334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCAGATACAGAAAATATCATGGCTCATGATG 393
OY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGluProHisSer 80
DB 394 GCRAAGCAACCATGATTTTCATGTTGTGAAAGAGAAATGATCAATGAGCCTCATTTCA 453
OY 81 AspArgIleTyrTyrLeuAlaMetSerGluLysAsnArgLysAsnThrLeuPheTyrSer 100
DB 454 GACAGATCTATTAATCTGTCATGTCGTGAGAAACAGAAATATCATCTGTTTATTTCT 513
OY 101 GluIleProLysThrIleAsnArgAlaValLeuMetLeuSerTyrLysProLeuLeu 120
DB 514 GAAATTCCTCAAACTATCATATGACGACAGCTTATGCTCTTGGAGCGCTTTTGG 573
OY 121 AspLeuPheGluIleThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuArg 140
DB 574 GATCTTTTTCAGCAACATGAGCTATGAGATGATATTCGAGAAAGAACTATTAGAG 633

OY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGluGlySerGly 160
DB 634 GAAAGAAAACGATTTGAAACAGTCGAAATGCTTTCTTACGATTTATACAGAGAGTGGCA 693
OY 161 ThrPheLeuPheGluIleGlySerGlyIleTyrHisValLysAspGlyGlyProGluGly 180
DB 694 ACATTTCTGTTTCACACCGGTAGTGAATTTATCATCTAAATAAGATGAGGCCCAAGAGA 753
OY 181 PheThrGluGluProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 754 TTTACGCAACACCTTTAAGGCCCATCTAGTGAAACTAGTTGCTCCACATACGATG 813
OY 201 AspProLysLeuGluProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 814 GATCCAAATTAATGCTCCCTGATTCAGACTGAGATGCTTTTATACATAGCAAGATATT 873
OY 221 TrpIleSerAsnIleValThrArgGluGluArgGluLeuThrTyrValHisAsnGluLeu 240
DB 874 TGGATATCTAACATCGTACACAGAAAGAAAGAGATCTATGTCACCAATGAGCTA 933
OY 241 AlaAsnMetGluGluAspAlaArgSerArgIleValAlaThrPheValLeuGluGlu 260
DB 934 GCCAACATGGAAGAGATGCGAATCAGCTGAGAGTCCCTACCTTGTCTCCAGAGAA 993
OY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280
DB 994 TTTGATGATATTCCTGCTATTTGGTGTGTCACAAAGCTGAAACATCTCCAGTGGTGGT 1053
OY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB 1054 AAAATTTCTTAAGAAATCTATATGAAAGAAATGATGATGAGGAGAAATATTATCATGTT 1113
OY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB 1114 ACATTCCTCATGTTGAAACAAAGAGGCGAGATTCATTCCTTATCTTAAACAGATGCA 1173
OY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 1174 GCAAACTCTAAAGTCACTTTTAAAGTGCACAAATAAGATTTGAGCTGAGAGAGATC 1233
OY 341 IleAspValIleAspLysGluLeuIleGluProPheGluIleLeuPheGluGluValGlu 360
DB 1234 ATGATGTCATATGATTAAGAACTAATTTCACTTTTGAGATTCTATTTGAAAGAGTTGAA 1293
OY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
DB 1294 TATATTTGCCAGAGCTGGATGAGCTCTGAGGAAATATGCTGGTCCATCTTACTAGAT 1353
OY 381 ArgSerGluThrArgLeuGluIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1354 CGCTCCCAAGACTCGCTACAGATGATGATCTCACTGAAATTTATTTATCCCACTAGAA 1413
OY 401 AspAspValMetGluArgGluAsnArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1414 GATGATGTTATGAAAGGACAGAGATCTATGATGATGATGATCTGATCTGAGCGCACATA 1473
OY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
DB 1474 ATTATATTAAGAAACAAACAGACATCTGATTAATATCATGACATCTTCACTGATTTT 1533
OY 441 ProGluSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
DB 1534 CCCCAAGTCAACAAAGAAATGAGTTATTTTGGCTTGAAATGCAAAACAGGTTTC 1593
OY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
DB 1594 CGTCATTTATACAAATTTATCATCTATTTTAAAGAAAGCAAAATTAACATGATCAGTGGT 1653
OY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB 1654 GGGCTGCTGCTCCAAAGTATTTCAAGTGCCTATCAAGAGAGAGATAGCAATTAACAGT 1713
OY 501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGluIleAlaAspGluValArgArg 520

```

Db      1714 GGTGAATGGAGAGTTCTTGCGCGCATGATCTAATATCCAAAGTTGAGGAAGTGAAGG 1773
Qy      521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSer 540
Db      1774 CTGGTATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGATGATGATGATGAT 1833
Qy      541 TyrValAsnProGlyValValThrArgLeuThrAspArgGlyTyrSerHisSerCysGly 560
Db      1834 TACGTAATCTCGAGAGGTACAAAGGCTGACTGACCTGGCTACTCATCTTCTTGCTGC 1893
Qy      561 IleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsnProHisCys 580
Db      1894 ATCAGTCAGCAGCTGATCTTTATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1953
Qy      581 ValSerLeuTyrLysLeuSerSerProGluLysAspProThrCysLysThrLysGluPhe 600
Db      1954 GTGTCCTTTTCAAGCTATCAAGCTCTGAAAGTGAAGCCCAAGCTTGCAGAAACAAAGAAATTT 2013
Qy      601 ThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620
Db      2014 TGGGCCACCATTTGGATTGACT----- 2036
Qy      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      2036 ----- 2036
Qy      641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db      2037 -----CCTCAGGTGCGAGTTG 2051
Qy      661 ValAsnAsnArgPheLysGlyValIleTyrPheArgLeuAsnThrLeuIleSerLeuGly 680
Db      2052 GTGAATATACGGTTTAAAGAGTCAAGTATTTCCGCTGAATACCTTGAAGCTCTCTAGGT 2111
Qy      681 TyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db      2112 TATGTGGTTGATGATAGACAAAGGGATCCTGTACCGAGGGCTTAAATTTGAAGGC 2171
Qy      701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db      2172 GCCCTTAAATATTAATGGGTCAAAATGAAATTCGATCGATGAGGAGGAGCTCCAAATAT 2231
Qy      721 LeuAspSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
Db      2232 CTACCTCTCGATATGATTTATGACTTAAGTGTGGGCATTCACGGCTGGGCTGCTAT 2291
Qy      741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db      2292 GGAAGATATACCTCTCCCTGATGGCATTTATGCAGAGGTCAATATCTTCAGGGCTTCATTT 2351
Qy      761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db      2352 GCTGGGGGCCAGACACTCTGTGATCTTATGATACAGATACACGGAACGTTATATATG 2411
Qy      781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db      2412 GGTACACCTGACCAAGATGAACAGGGCTATTACTTAGATCTGGGCCATCAAGCAAGAA 2471
Qy      801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2472 AAGTTCCCTCTGAACCAATCGTTTACTGCTTTACATGGTTTCTCGAGTGAAGAAATGTC 2531
Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyrAsp 840
Db      2532 CATTTTCACATACACAGTATATTACTGAGTTTATAGAGGCTGGGAAACCATATGAT 2591
Qy      841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db      2592 TTACAGATCTATCTTCAGAGAGACACAGCATTAAGAGTTCTCGAATCGGGAACATATAT 2651
Qy      861 GluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880

```

```

Db      2652 GAAGTCATCTTTTGACACTTACACTTCAAGAAAACCTTGATACAGTATGCTGCTTAAAA 2711
Qy      881 ValIle 882
Db      2712 GTGATA 2717
RESULT 10
ABK83331
ID ABK83331 standard; cDNA: 4676 BP.
XX
AC ABK83331;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-1 splice variant #7.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO20021134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
XX
DR P-PSDB; ABG61600.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS Disclosure; Page 72-73; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypertension,
CC hyperextension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0
Score: 4385.00
Percent Similarity: 94.22%
Best Local Similarity: 94.22%
Query Match: 93.30%
DB: 24
Length: 4676
Matches: 831
Conservative: 0
Mismatch: 1
Indels: 51
Gaps: 1

```


US-10-070-464-1 (1-882) x ABK83331 (1-4676)

QY 1 MetAlaAlaMeGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
 Db 214 ATGGAGCGCAGCATGGAAAACAGAACGCTGGGTGGAGATATTTGAACTCGGACTGCT 273
 QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 Db 274 GAGGAGAAATTTGAATTCACAGATCGGCTTAAATTTGAGCTTTTATTTATGTGAGCGGTAT 333
 QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
 Db 334 TCTCGAGTCAGCTTAAAAAGCTGCTTCCGATACCAAGAAAATATCATGCTCATCATGATG 393
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
 Db 394 GCTAAGGACCAACATGATTCATGTTTGGAAAGAGAAATGATCCAGATGAGCTCATTTCA 453
 QY 81 AsparGlyIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
 Db 454 GACAGAAATCTATTCCTTCCATGTGTGTGAGAACAGAAATACACTGTTTATTTCT 513
 QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
 Db 514 GAAATTTCCAAAACATATCATATGAGACGAGCTTATATGCTCTCTTGGAAAGCTCTTTTG 573
 QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuArg 140
 Db 574 GATCTTTTTCAGGCAACCTGGACTATGGAATGTATTTCTGAGAAAGAAACACTATTAA 633
 QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnLysSerGly 160
 Db 634 GAAAGAAAACGATTTGGAACAGTCGGAATTTGCTTCTTACGATTTATACCAAGAAAGTGA 693
 QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
 Db 694 ACATTTCTGTTCAGACCGGTAGTGGAATTTATCATGTAAGATGAGAGGCGACACAGGA 753
 QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
 Db 754 TTTAGCCAAACACTTTTAAAGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGA 813
 QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
 Db 814 GATCCAAAATTTATGCCCTGCTGATCCAGACTGGATTGCTTTATATCATAGCAACGATAT 873
 QY 221 TyrIleSerAsnIleValIleThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
 Db 874 TCGATATCTAAACATCGTAACAGAGAAAGAGAGACTCATATGTGCAATGAGCTA 933
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
 Db 934 GCCAACATGGAAGAAAGATCCAGATCCAGTGGAGTGCATCCCTTGTCTCCAAAGAA 993
 QY 261 PheAspArgTyrSerGlyTyrTyrTyrCysProLysAlaGluThrThrProSerGlyGly 280
 Db 994 TTTGATAGATATTTGGCTATTCGTGTGTCCAAAAGCTGAAACAACTCCAGTGGTGT 1053
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
 Db 1054 AAAATTTCTTGAATCTATATGAAAGAAATGATGATCTGAGGTGGAATTTATTCATGTT 1113
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 1114 ACATCCCTCATTTTGGAAACAGAGAGGCGAGATTCATTCCTTATCTTAAACAGGTACA 1173
 QY 321 AlaAsnProLysValIleThrPheLysMetSerGluIleMetIleAspAlaGluIleArgIle 340
 Db 1174 GCAAAATCCTAAGCTCATTTTAAAGATGTGAGAAATATGATGATGCTGAAGAAAGATC 1233
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
 Db 1234 ATAGATGTATGATTAAGAACTAATTCACCTTTTGAATTCATTTGAAAGAGATTGA 1293

QY 361 TyrIleAlaArgAlaGlyThrThrProGluGluLysLysTyrAlaTyrSerIleLeuLeuAsp 380
 Db 1294 TATATTTCCAGAGCTGAGTGAAGCTCTGAGGAAAAATATGCTTGGTCCATCCACTAGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlnIleValIleuIleSerProGluLeuPheIleProValGlu 400
 Db 1354 CGCTCCAGACTCGCTACAGATAGTGTGATCTCACTCGAATTTATTTATCCAGTAGAA 1413
 QY 401 AsparValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTATGGAAGGACAGAGACTCATTTGAGTGCAGTCCGATCTCTGACGCCACTA 1473
 QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
 Db 1474 ATTATCTATGAAGAAACACAGACATCTGGAATTAATTCATGACATCTTTCAATGTTTT 1533
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 Db 1534 CCCCAAGTCAGAAAGAGAAATTTGAGTTTATTTTGGCTCGAATGCAAAACAGCTTTC 1593
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerGly 480
 Db 1594 CGTCATTTATACAAAATTTACATCTATTTTAAAGAAAGCAAAATATTAACGATCCAGTGT 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 Db 1654 GGGCTGCTGCTCCAGAGATTCAGAGTCTCTATCAAAAGAGAGATACCAATTTACAGT 1713
 QY 501 GlyIleTyrGluValLeuGlyValArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 Db 1714 GGTGAATGGAAAGTTCTTGGCCGCGATGATGATTAATTCAGATTGATGAATCGAAG 1773
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSer 540
 Db 1774 CTGGTATATTTTGAAGGACCAAAAGACTCCCTTATGAGCATCACCTGTACACTAGTCACT 1833
 QY 541 TyrValAsnProGlyGluValIleThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
 Db 1834 TAGGTAATCTCTGAGAGGTGCAAGGCTGAGTCAAGCGGTACTACACTTGTGCTGC 1893
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
 Db 1894 ATCAGTCAGCACTGAGACTCTTTATTAAGTAAGTATGTAACAGAAATCCACACTGT 1953
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 Db 1954 GTGTCCCTTTACAAAGCTATCAAGTCTGAGAGATGACCCAACTTGCAAAAACAAAGGATTT 2013
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
 Db 2014 TGGGCCAACATTTGATTCAGACAGGCTCTCTCTGACTATATCTCCAGAAATTTTC 2073
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 Db 2074 TCTTTTGAAGACTACAGTGAATTTACATTTGATGAGAGTCTTACAGCTCATATGATCA 2133
 QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
 Db 2134 CAGCTCGAAGAAATATCTACTGTGCTTCAATATATGGTGTGCG----- 2180
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 Db 2180 ----- 2180
 QY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 Db 2180 ----- 2180
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
 Db 2181 -----GGTCAATATGAAATTTGACGATCGATCGTGAAGGACTCCCAATAT 2222

QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyr 740
DB 2223 CTACCTTCTCATATGATGATTTTATGACTTACATGCTGGGACATCCAGCGCTGCTCAT 2282
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2283 GAGAGATACCTCTCCCTGATGCGATTAAATGACAGAGTCAGATATCTTACGGGTTGCTATT 2342
QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
DB 2343 GCTGGGGCCCAAGTACCTCTGATCTTCTATGATACAGATACACGGAACGTTATATG 2402
QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
DB 2403 GGTACCTGACCAAGAAAGAAACAGGGCTATTACTTAGATCTGGCCATGCAAGCAGAA 2462
QY 801 LysBheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
DB 2463 AAGTTCCCTCTGAACCAATCGTTTACTGCTTACATGGTTCTCGATGAGATGTC 2522
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
DB 2523 CATTTTGACATACACAGATATTACTGAGTTTGTAGAGGGCTGGAAACCATATGAT 2582
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlySerGlyGluHisTyr 860
DB 2583 TTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTCTGAATCGGAGAACATTAT 2642
QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880
DB 2643 GAATGCAATCTTTTGCACTACTTCAAGAAACCTTGATCAGATATTCGCTCTAATA 2702
QY 881 ValIle 882
DB 2703 GTGATA 2708

RESULT 11
ABNS9774
ID ABNS9774 standard; cDNA, 2842 BP.
AC ABNS9774;
XX 28-JUN-2002 (first entry)
XX Novel human coding sequence SEQ ID NO: 185.
DE Human; anti-nausea; vulnerability; anti-inflammatory; immunomodulator;
KW anti-infectivity; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
KM neuroprotective; anti-parkinsonian; protein therapy; EST;
XX expressed sequence tag; gene; ss.
OS Homo sapiens.
XX MO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001MO-US26015.
XX 11-SEP-2000; 2000US-0659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX P-PSDB; AB897361.
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -
XX Claim 1; SEQ ID NO 185; 509BP; English.

XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haemacopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
XX invention.
SO Sequence 2842 BP; 857 A; 592 C; 635 G; 758 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 2842
Score: 4118.00 Matches: 782
Percent Similarity: 88.66% Conservative: 0
Best Local Similarity: 88.66% Mismatch: 0
Query Match: 87.62% Indels: 100
DB: 24 Gaps: 1

US-10-070-464-1 (1-882) x ABNS9774 (1-2842)

QY 1 MetAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
DB 234 ATGCACACAGCAATGAAAGCAAGACAGCTGGGTGAGATATTGAAACTGCGGACTGT 293
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
DB 294 GAGGAGATATGATGATACAGGATCGGCTTAATTTGAGCCCTTTTGTGGACCGGTAT 353
QY 41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
DB 354 TCCGAGAGTCAGCTTAATAAGCTGCTTCCGATACCAAGAAATATCATCGCTATATGATG 413
QY 61 AlaLysAlaProHisAspPheMetPheValIleArgAsnAspProAspGlyProHisSer 80
DB 414 GCTAAGCACACACATGATTTTATGTTGTAAGAGAAATGATCCAGATGACCTCATTTCA 473
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 474 GACGAATCTATTAATCTTGCATGCTGTGTAGAAACAGAAATATACATGTTTATTTCT 533
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
DB 534 GAATTCACCAAACTATCAATAGACAGCATCTTAATGCTCTTGGAAACCTCTTTTG 593
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB 594 GATCTTTTTCAGCAACACTGACCTAGATGATATTCTCGAAGAAAGAACTATTAGA 653
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 654 GAAAGAAAGCAATGGAACAGTCGGAATTCCTTTTACATTAATCAACAAAGAGTGA 713
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
DB 714 ACATTTCTGTTTCAAGCCGTAAGTATGATTAATCACTTAAGATGAGGACCAAGGA 773
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 774 TTACGCAACAACTTTAAGGCCCAATCTAGTGAATCTAGTTGCCAATATCGGATG 833
QY 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
DB 834 GATCCAAATTAATGACCTCTGATCCAGATGATGCTTTTATACATACCAAGATATT 893
QY 221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
DB 894 TGGATATCTTAACATCGTAACCAAGAAAGAGAGCTCATTAATGCAATAGACTA 953

QY 241 AlaasmeGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluGlu
 Db 954 GCCACATCGAAGAAAGATGCCAGATCAGCTGAGCTCGTACTTGTCTCCAGAAAGAA
 QY 261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly
 Db 1014 TTTCATAGATATCTCGCTATGCTGATGCTCAAAAGCTGAAACATCCAGAGTGCT
 QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal
 Db 1074 AAAATTCTGAAATTCATATAGAAAGAAATGATCAATCTGAGGTGAATTAATTCATGTT
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr
 Db 1134 ACATCCCTATGTTGAAACAGAGAGGAGGAGATTCCTCCGTTTCTTAAACAGGTACA
 QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle
 Db 1194 GCAATCTCTAAAGTCACTTTTAAAGATGTCAGAAATTAATGATGATGCTGAAGAGATC
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu
 Db 1254 ATGAGTGTCAATGATTAAGAACTAATTCACCTTTTGAGATTCATTTGAAGAGATTGAA
 QY 361 TyrIleAlaArgAlaGlyTyrTrpThrProGluGlyLysTyrAlaTyrPheSerIleLeuLeuAsp
 Db 1314 TATATTCACAGAGCTGATGAGCTCTGAGGAGAAATATGCTTGCTTCATCTCACTAGAT
 QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu
 Db 1374 CGCTCCAGACTCCCTCAAGATAGTGTGATCTCACCTGAATTAATTAATCCAGTAGAA
 QY 401 AspAspValMetGluValArgGlnArgLeuIleGluSerValProAspSerValThrProLeu
 Db 1434 GATGATGTATGGAAGACAGAGACTCATGATGACGCTGATTCGTGAGCCCACTA
 QY 421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe
 Db 1494 ATATATCATGAAGAAACAAGACATCGATTAATTCATGACATCTTTCATGCTTTT
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe
 Db 1554 CCCCAAGTCAAGAAAGAGAAATGAGTTATTTTTCCTCTGAAATCAAAACAGTTTC
 QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly
 Db 1614 CGTATTTATACAAATTAATCATCTAATTTAAAGAAAGCAAAATTAACGATCCAGTGT
 QY 481 GlyLeuProAlaProSerAspPheLysCysArgProIleLysGluGluIleAlaIleThrSer
 Db 1674 GGGCTGCTGCTCCAAAGGATTTCAAGTCTCTATCAAGAGAGATAGCAATTAACAGT
 QY 501 GlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg
 Db 1734 GGTGAATGGGAAGTCTTGGCGGCGCATGATCTAATATCCAAAGTTGATGAAGTCAGAGG
 QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSer
 Db 1794 CTGGTATATTTGAAGGACCAAGACCTCCCTTTAGAGATCACCTGTAGTATGTCAGT
 QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys
 Db 1854 TACGTAAATCCTGAGAGGTGACAAAGCTGACCTGCTGCTACTCACAATTTTGTGCTGC
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerArgGlnLysAsnProHisCys
 Db 1914 ATCAGTCAGACCTGTGACTTTTATTAAGTAAGTAAAGTAAAGCAAGAAATCCACACTGT
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe
 Db 1974 GTGTCCTTTTACAAGCTATCAAGCTCTGAAGATACCACTTGCAAAACAAAGAAATTT
 QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe

Db 2034 TGGGCCACCATTTTGGATTGACGAGCTCTCTTCTCATACTACTCTCCAGAAATTTTC
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu
 Db 2094 TCTTTTAAAGTACTACTGATTTTACATGTATGAGATGCTTACAAAGCTCATGATCTA
 QY 648 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu
 Db 2154 CAGCTGGAAGAAATATCTCTAGCTGTCTTCAATATATGATGAGTCTCAG-----
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly
 Db 2204 -----
 QY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly
 Db 2204 -----
 QY 701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr
 Db 2204 -----
 QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyr
 Db 2204 -----
 QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle
 Db 2205 -----GTTGCTAATT 2213
 QY 761 AlaGlyValAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet
 Db 2214 GCTGGGGCCCAAGCTCTGTGATCTTATGATACAGAGATACAGAGAACTTATATG
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu
 Db 2274 GGTCACTGACCAAGATGAACAGGCTATTACTTAGATCTGTGGCCATGCCAAGCAAA
 QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal
 Db 2234 AAGTTCCTCTGAAACCAAACTCTTACTCTCTTACATGATGTTCCGTGATGAAATGTC
 QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp
 Db 2294 CATTTGCACATACCAATATATTAATGATTTTATGAGGCTGGAAGCCATATGAT
 QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyr
 Db 2454 TTACAGATCTATCTCTCAGAGAGACACAGCATAAAGATTCCTGAATCGGAGAACTATAT
 QY 861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys
 Db 2514 GAATCGATCTTTTGCATCACTTCAAGAAACCTTGATCAGATATGCTGCTGCTAAAA
 QY 881 ValIle 882
 Db 2574 GTGATA 2579
 RESULT 12
 ID ABK83325
 XX ABK83325 standard; cDNA; 4523 BP.
 XX
 AC ABK83325;
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPP-1 splice variant #1.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP-1;
 KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;

KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder; gene; ss.
 OS Homo sapiens.
 PN MO200231134-A2.
 XX 18-APR-2002.
 PD 12-OCT-2001; 2001MO-US31874.
 PF 12-OCT-2000; 2000US-240117P.
 PR 12-OCT-2000; 2000US-240117P.
 XX (FERR) FERRING BV.
 PA Qi S, Akinsanya KO, Riviere PJ, Junien J;
 PI MPI; 2002-444178/47.
 XX P-PSDB; ABC61594.
 DR New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 PS Disclosure; Page 61-62; 113pp; English.
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPP-IV proteins.
 CC
 SQ Sequence 4523 BP; 1384 A; 828 C; 940 G; 1371 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4523
 Score: 4092.50 Matches: 780
 Percent Similarity: 88.44% Conservative: 0
 Best Local Similarity: 88.44% Mismatches: 1
 Query Match: 87.07% Indels: 102
 DB: 24 Gaps: 1
 US-10-070-464-1 (1-882) x ABK83325 (1-4523)
 QY 1 MCTAALALAMETGLUTHRGUGLNLGULVAGLULPHEGLUTHRRLAASPQYS 20
 DB 214 ATGGCAGCAGCAATGAGAAACAGAACAGCTGGGCTTGATATTGAACTGGGACTGT 273
 QY 21 GUGLUAANLLEGLUSERGLNAPAPGPOLYLEUGLUPROBHEGYRVALGUATGYR 40
 DB 274 GAGGAGAAATATTGAATCAAGAGATCGGCTTAATTTGAGAGCTTTTATGTGAGCGGTAT 333
 QY 41 SERTPSEGLNLEULYLSLEULEUALAAPTTHRGYLYRHSGLYTYRMECTMEC 60
 DB 334 TCCCTGAGTCACTTAAAGCTGCTTGCGATGACCAAGAAATATATGCTTACATGAGTG 393
 QY 61 ALALYSALAPROHISAPRHEMECTPHEVALLYSARGASAPPROASPGLYPROHISER 80
 DB 394 GCTAAGCCACCATATTTTCAATGTTTGGAGAGAGAAATGATTCAGATGACCTCATTTCA 453
 QY 81 ASPAPGLILETYRTRYLEUALAMECTSERGLYGLUASARGLUASNTHRLEUPHERYSER 100

DB 454 GACAGAACTATTACCTTGGCATGTCTGTGAGAACAGAGAAAATACACTGTTTATTC 513
 QY 101 GULLEPROLYSTHRIEANAARGAIALAVALLEUWELTUSERTPLYPPOLEULEU 120
 DB 514 GAAATTCGCAAACTAATCAATAGAGCAGACTTTAATGCTCTCTTGGAAACCTCTTTTG 573
 QY 121 ASPLEUPHEGLNALATHRLEAAPTGYRGLYMETYRSEARAGGUGLUGLULEULEUARG 140
 DB 574 GATCTTTTCAAGGCACACTGACTGATATGAAATGATTTCTGAGAAAGAAACTTATTAGA 633
 QY 141 GLUARGLYSARGILEGLYTHRVALLGLYILEALASERTYRAPRYRHISGLINGLYSERGLY 160
 DB 634 GAAAGAAAACGATTGGAAACAGTCGGAATTCCTTACGATTATACCAAGAAAGGAGA 693
 QY 161 THRPHLEUPHEGLNALAGLYSERGLYILETYRHIVALLYSAPGLYGLYPROGLINGLY 180
 DB 694 ACATTTCTGTTTAAAGCCGTAGTGGAAATTTATCACTTAAAGATGAGGCCCAAGGA 753
 QY 181 PHETHRGLINGLNPPOLEUAARGPROASNTLEUVAGLUTHRSECYSPROASNTLEARGMET 200
 DB 754 TTTACGCAACACTTTTAAGGCCCAATCTAGTGAATCTGTTGCTCCACATACGAGATG 813
 QY 201 ASPPTLYLEUCYSEPROALASPPTSPTRILEALAPHEILEHISERASNSAPILE 220
 DB 814 GATCCAAATATTAGCCCTGCTGATCCAGACTGGATTCCTTTATACATACGAACGATATT 873
 QY 221 TTPILSEERANLLEVALTHRVARGLUGLUAARGLEUTHRYRVALHISANGLULEU 240
 DB 874 TGGATATCTTAACCTGTAACAGAGAAAGAAAGAGCTCACTTATGTCACCAATGAGCTTA 933
 QY 241 ALASAMEGLUGLUNAPALARGSERVALGLVALLAPHRPHEVALLEUGLUGLUGLU 260
 DB 934 GCCAACGTGAAAGATGACAGTCACTGAGAGTGCCTACTTGTCTCCAAAGAA 993
 QY 261 PHEASPARGYRSEGLYTRYTPTRCYSPROLYSALAGLUTHRTPROSEGLYGLY 280
 DB 994 TTTGATAGATATCTGGCTATGTGGTGTGCCAAAAGCTGAACCACTCCACAGTGGT 1053
 QY 281 LYSILELEUARGLILEUPTYRGLUGLUAASAPGLUSERGLVAGLULLEHISVAL 300
 DB 1054 AAAATTTCTTAAGAAATCTATATGAAAGAAATGAGAAATCTGAGTGGAAATTAATTCATGTT 1113
 QY 301 THRSERPROMETLEUGLUTHRARGARGALASPSEPRHEARGTRYRPOLYSTHRYTHR 320
 DB 1114 ACATCCCTTATGTGGAAACAGAGAGGAGATTCAATCCGTTATCTTAAACAGGTACA 1173
 QY 321 ALASAPPROLYSEVALTHRPHLYSMETSESGULILEMETILEASPALAGLUGLYARGILE 340
 DB 1174 GCAAATCTTAAGATCACTTTTAAAGATGACGAAATATATGATGATGCTGAAGAGAGATC 1233
 QY 341 ILEASPVALLLEASPLYSGLULEULLEGLNPROPHGLULLEUPHEGLUGLYVALGLU 360
 DB 1234 ATAGATGATATGATATAGAAACCTAATTCACCTTTTGAAGTCTTATTTGAAGAGAGTTGAA 1293
 QY 361 TYRILEALARGALAGLYTRTPRPROGLUGLUVLYSTYRALSTPSERIILEULEUASP 380
 DB 1294 TATATTGCGAGAGCTGATGACTCTGAGGAAATATGCTTGTCACATCTTACATGAT 1353
 QY 381 ARGSEGLNTHRARGLEUGLULLEVALLEULELSESPROGLULEUPHEILEPROVALGLU 400
 DB 1354 CGCTOCCACACTCGCTTACAGATAGTGTGATCTACCTGAAATATTATCCACAGTAGAA 1413
 QY 401 ASPAPSVALLMEGLUARGLUGLUGLULEULLEGLUSEVALPROASPSEVALTHRPOLEU 420
 DB 1414 GATGATGTTATGAAAGGACAGAGACTCATGTGATGCTGCTGATTCCTGAGCGCACGTA 1473
 QY 421 ILEILETYRGLUGLUTHRTHRAPLLETTPRIEASNTLHISASPTLLEPHEHISVALPHE 440
 DB 1474 ATTAATCTATGAAGAAACAGACATCTGATTAATATCAAGACATCTTTCATGTTT 1533
 QY 441 PROGLINSEHILEGLUGLUGLULLEGLUPHEILEPHEALSESGLYUCYSYSTRGLYPHE 460
 DB 1534 CCCCAAGTCAGAGAGAGAAATGAGTTATTTTGGCTCTGAAATGCAAAACAGGTTTC 1593

Oy	461	ArgHisLeuTYrLysIleThiSerIleLeuLysGluSerLysTYrLysArgSerSerGly	480
Db	1594	CGTCATTATATACAAATATTACATCTATTTTAAAGGAAGCAATATAAACGATCCAGTGGT	1653
Oy	461	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThiSer	500
Db	1654	GGGCGCCCGCTCCAAAGTATTTCAAGTGCTCTATCCAAAGAGAAATGCAATTACCGAGT	1713
Oy	501	GlyGluTPGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg	520
Db	1714	GGTGAATGGGAAGATTCTTGCCGGCATGGATTTAATATCCAAAGTTGATGAAGTCCGAAG	1773
Oy	521	LeuValIYrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTYrValAlaSer	540
Db	1774	CTGGATATATTTTGAAGGCGCACCAAAAGCTCCCTTTAAGAGCATCACCTGTACGTAGTCA	1833
Oy	541	TYrValAsnProGlyGluValIleThrArgLeuThrAspArgGlyTYrSerHisSerCysCys	560
Db	1834	TACGTAATATCCGTGGAGAGGTGCACAAAGGCTGACCTGACCGGTGATCTCAATTCTGCTGC	1893
Oy	561	IleSerGlnHisCysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisCys	580
Db	1894	ATCAGTCAGCACTGTGACCTTTTATAGTAAGTATAGTAACCAAGAAATCCACACTGT	1953
Oy	581	ValSerLeuTYrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe	600
Db	1954	GTGTCCCTTTACAGCTATCCAAAGTCCCTGAAAGATGACCCCAACTGGCAAAACAAAGCAATTT	2013
Oy	601	TPAlaThrIleLeuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePhe	620
Db	2014	TGGGCCACCATTTTGGATTCCACAGGTCTCTTCTGATATACCTCCCGCAAAATTTTC	2073
Oy	621	SerPheGluSerThrThryPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeu	640
Db	2074	TCTTTTGAAGAGTACTACTGATTTACATGTATGGGATGCTCTACAGCTCATGATCTTA	2133
Oy	641	GlnProGlyLysLysTYrProThrValLeuPheIleTYrGlyGlyProGlnValGlnLeu	660
Db	2134	CAGCTCGGAAACAAATATCTACTGCTGCTCATATATGGGGT-----	2178
Oy	661	ValAsnAsnArgPheLysGlyValIleTYrPheArgLeuAsnThrLeuAlaSerLeuGly	680
Db	2178	-----	2178
Oy	681	TYrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly	700
Db	2178	-----	2178
Oy	701	AlaPheLysTYrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYr	720
Db	2178	-----	2178
Oy	721	LeuAlaSerArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrSerTYr	740
Db	2178	-----	2178
Oy	741	GlyGlyTYrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle	760
Db	2179	-----CGGTT-GCTATT	2188
Oy	761	AlaGlyAlaProValThrLeuThrIlePheTYrAspThrGlyTYrThrGluAspGlyMet	780
Db	2190	GCTGGGGCCCCAGTCACTGTGGATCTTATAGTATACAGGATTAACGGAACGTTATATG	2249
Oy	781	GlyHisProAspGlnAsnGluGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlu	800
Db	2250	GGTCAACCCCTGACCAATGAACAGGAGCTATTACTTATAGATCTGTGGCCATGGAACAGAA	2309
Oy	801	LysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal	820
Db	2310	AAGTTCCTCTGAAACCAATGTTTACGTCTTACATATGGTTTCTCGATGAGAAATGTC	2368

Qy	821	HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaAlaIleLysProTyrAsp	840
Db	2370	CATTTCGACATCAACAGATATATTACTGAACTTTTAAAGTAGGGCTGGAAAGCCATATGAT	2422
Qy	841	LeuGlnIleTyrProGlnIuArgHisSerIleArgValProGlnuSerGlyGlnHisTyr	866
Db	2430	TTTACGATCTATCTCCAGAGAGACACACGATTAAGAGTTCCTGAAATCCGGAGAAACATTAT	2488
Qy	861	GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys	880
Db	2490	GAACTGCATCTTTTGCACACTACTCTCAAGAAACCTTGATCAAGTATTCCTGCTTAATAA	2544
Qy	881	ValIle 882	
Db	2550	GTGATTA 2555	
RESULT 13			
AAD23843			
ID	AAD23843	strand:rd; cDNA; 2510 BP.	
XX	AAD23843;		
AC			
XX			
DT	07-MAR-2002	(first entry)	
XX			
DE	Human protease PRTS-2 cDNA.		
KW	Human; protease; PRTS-2; tranquilizer; gene therapy; vaccine; allergy;		
KW	infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;		
KW	atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;		
KW	gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;		
KW	epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;		
KW	hypertension; neurological disorder; Parkinson's disease; drug screening;		
KW	cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;		
KW	diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;		
KW	autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;		
KW	developmental disorder; reproductive disorder; infertility; diarrhoea;		
KW	dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.		
XX			
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	616..2358	
FT		/tag= a	
FT		/product= "Human protease PRTS-2 protein"	
XX			
PN	WO200183775-A2.		
XX			
PD	08-NOV-2001.		
XX			
XX	04-MAY-2001; 2001WO-US14651.		
XX			
PR	04-MAY-2000; 2000US-202082P.		
PR	11-MAY-2000; 2000US-203566P.		
PR	17-MAY-2000; 2000US-205803P.		
PR	25-MAY-2000; 2000US-207477P.		
PR	01-JUN-2000; 2000US-209402P.		
XX			
XX	(INCYTE) GENOMICS INC.		
PA			
PI	Deleage AM, Lai P, Hafalia A, Patterson C, Wallia NK, Kearney L;		
PI	Trilbouley CM, Khan PA, Yao MG, Baughn MR, Aizawa Y, Elliott VS;		
PI	Nguyen DB, Gandhi AB, Yang J, Hernandez R, Policky JL, Lu DM;		
PI	Reddy R, Yue H, Tang YT;		
XX			
XX	WPI; 2002-034518/04.		
DR	P-PSDB; AAE14337.		
XX			
PT	Novel human proteases and polynucleotides encoding the proteases,		
PT	useful for treating, diagnosing or preventing cell proliferative,		
PT	cardiovascular, autoimmune/inflammatory, neurological and developmental		
PT	disorders -		
XX			
PS	Claim 5; Page 139-140; 151PD; English.		


```
Oy 569 eSerIySTySerAnGlnLyAsnProHicYsValSerLeuTyLyLysLeuSerSerPr 589
Db 1581 AAGTAAGTAAGTAACCAAGAAATCCACACTGTGTCCCTTACCAAGACTACAGTCC 1640
Oy 589 OGluAspAspProThrCysLySthLySgluPheTrpAlaThrIleLeuAspSerAlaGI 609
Db 1641 TGAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACCACTTTTGGATTGACGACGG 1700
Oy 609 YProLeuProAspTyThrProProGlnIlePheSerPheGluSerThrThrGlyPheTh 629
Db 1701 TCCCTTCTCTACATACATACCTCCAGAAATTTCTCTTTTGAAGTACTACTGATTTTAC 1760
Oy 629 rLeuTyGlyMetLeuTyLySProHicAspLeuGlnProGlyLySlySTyProThrVa 649
Db 1761 ATTGTATGCGATGCTCTCAAGACCTCAAGATCTACAGCTGAAAGAAATATCTACTGT 1820
Oy 649 lLeuPheIleTyGlyGlyProGlnValGlnLeuValAsnAsnArgPheLySglYValLy 669
Db 1821 GCTGTTCATATATGTGTGCTCCAGGTGCAGTGTGTAATATCGGTTTAAAGGAGTCAA 1880
Oy 669 sTyTPheArgLeuAnthrLeuAlaSerLeuGlyTyTrpAlaValValIleAspAsnAr 689
Db 1881 GTATTTCCGCTTGAAATACCTCAAGCTCTCTAGATTATGTGTGTAGTATGACAAACAG 1940
Oy 689 gGlySerCysHisArgGlyLeuLySPhelGlyValaPheLySTyLySmetGlyGlnI 709
Db 1941 GCGATCCGTCACCGAGGCGCTTAAATTTGAAAGGCGCTTAAATTAATAAG----- 1992
Oy 709 eGluIlleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPheIleAs 729
Db 1992 ----- 1992
Oy 729 pLeuAspArgValGlyIleHisGlyTrpSerTyGlyTyTrleuSerLeuMetAlaLe 749
Db 1992 ----- 1992
Oy 749 uMetGlnArgSerAspIlePheArgValAlaIleAlaGlyValaProValThrLeuTrpI 769
Db 1993 -----GTTGCTATTGCTGGGCGCCCACTGCTGTGAT 2027
Oy 769 ePheTyAspThrGlyTyTrThrGluArgTyMetGlyYHisProAspGlnAsnGluGln 789
Db 2028 CTTCTATGATACAGAGATACAGGAACGTTATATGGGTACACCTTACCAAGATGAACAGGG 2087
Oy 789 YTyTyTyLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArgLe 809
Db 2088 CTATTACTTAAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCTCTGAAACCAAAATCGTT 2147
Oy 809 uLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 829
Db 2148 ACTGCTCTTACATGCTTCCCTGAGTAGAATGTCATTTTGCACATACCAATATATTACT 2207
Oy 829 uSerPheLeuValArgAlaGlyLySProTyAspLeuGlnIleTyProGlnGluArgHis 849
Db 2208 GAGTTTATTTAGTGGGCTGGAAAGCCATATAGATTTA-----CAGGAGAACACA 2255
Oy 849 sSerIleArgValProGluSerGlyGluHisTyGlyuLeuHisLeuLeuHisTyTrleuG 869
Db 2256 CAGCATTAAGACTTCTGAATCGGAGAAACATTAATGAATCGATTTTGGACCTACCTTCA 2315
Oy 869 nGluAsnLeuGlySerArgIleAlaAlaLeuLySValIle 882
Db 2316 AGAAAACTTGATCAGTATTCCTGCTTAAAGTATTA 2355
```

RESULT 14

ABN59775 standard; cDNA; 2668 BP.

AC ABN50775;

DT 28-JUN-2002 (first entry)

XX Novel human coding sequence SEQ ID NO: 186.

```
XX Human; antihaemic; vulnery; antiinflammatory; immunomodulator;
KW antileukemia; cerebroprotective; cytosolic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
OS Homo sapiens.
PN WO200222660-A2.
PD 21-MAR-2002.
PE 10-SEP-2001; 2001WO-US26015.
PR 11-SEP-2000; 2000US-0659671.
PA (HYSE-) HYSEQ INC.
PI Tang YF, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
DR WPI; 2002-292408/33.
DR P-PSDB; ABB97362.
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -
XX Claim 1; SEQ ID NO 186; 509pp; English.
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haemopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention.
XX SQ Sequence 2668 BP; 796 A; 564 C; 592 G; 716 T; 0 other;
XX
XX Alignment Scores:
XX Score: 0 Length: 2668
XX Percent Similarity: 3771.00 Matches: 724
XX Best Local Similarity: 82.09% Conservative: 0
XX Query Match: 80.23% Mismatches: 0
XX DB: 24 Gaps: 158
XX
XX US-10-070-464-1 (1-882) x ABN59775 (1-2668)
Oy 1 MetAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 234 ATGGCAGCAGCAATGAGAAACAGAACAGCTGGGTGTATGATTTGAAATCGCGGACTGT 293
Oy 21 GluGluAsnIleGluSerGlnAspArgProLySLeuGluProPheTyValGluArgTy 40
Db 294 GAGAGCAATATTAATACAGAGATCGGCTTAAATTTGAGAGCTTTATGTTGACGGTAT 353
Oy 41 SerTrpSerIleuLySlySLeuLeuAlaAspThrArgTyGlyYHisGlyTyTrMetMet 60
Db 354 TCCGTGAGTCAAGCTTAAAGAGCTGCTGCGATACCAAGAAATATCATGCTCATGATG 413
Oy 61 AlalyeAlaProHisAspPheMetPheValIySArgAsnAspProAspGlyProHisSer 80
Db 414 GCTAAGCAGCAGATATGATTTTCAATGTTTGAAGAGAAATATCAAGATGACCTCATTC 473
Oy 81 AspArgIleTyTyTrleuAlaMetSerGlyGluAsnArgGluAnthrLeuPheTySer 100
Db 474 GACAGATATTAATACCTTGCATGTCTGGTGAAGAACAGAAATATCACTGTTTATTC 533
```

QY 101 GluIleProLysThrIleAsnArgAlaIleValLeuMetLeuSerTrpLysProLeuLeu 120
DB 534 GAATTCCTCCAAACCTATCATATGACACAGCTTTAAGCTCTCTTGGAAAGCTCTTTTG 593
QY 121 AspheuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB 594 GATCTTTTTCAG----- 605
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 605 ----- 605
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
DB 605 ----- 605
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 606 -----CAACAACCTTTAAGGCCCAATCTAGTGGAACTAGTGTCCCAACATACGATG 659
QY 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
DB 660 GATCCAAATTAATGCTCGCTGATCCAGACTGGATTCCTTTATACATAGCAACGATATT 719
QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
DB 720 TGGATATTTACATCGTACCAAGAAAGAAAGAAAGAACTCACTTATGTCCAAATAGACTA 779
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
DB 780 GCCAAACATGAAGAAGATGCCAGATCAGCTGAGTCCGACTTGTCTCCAAAGAA 839
QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGluThrThrProSerGlyGly 280
DB 840 TTTTATGATATTTCGCTATTGCTGTGTCCAAAGCTGAACAACACTCCACAGTGTGT 899
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB 900 AAAATTCTTACAAATTCATATGAAACAAATGATGATGTGAGTGGAAATTAATTCATTT 959
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB 960 ACATCCCTATGTTGGAAACAAGAGGCGCAGATTCATTCGGTTATCTTAAACAGGTGCA 1019
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlnGlyArgIle 340
DB 1020 GCAAAATCTTAAGTCACTTTAAGATGTCAAGAAATGAATGATGCTGAAGGAAGGATC 1079
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGlnIleLeuPheGluGlyValGlu 360
DB 1080 ATGATGTCAATACATACAACTAATTCACCTTTTGAGATTCATATTGAAAGAGTTGAA 1139
QY 361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
DB 1140 TATATTGGCAGAGCTGATGAGTCTCTGAGGAAATATGCTTGATCCATCTACTAGAT 1199
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1200 CGCTCCCAAGACTCGCTCAATAGTGTGATCTCACCTGAATTAATTATCCAGTAA 1259
QY 401 AspAspValMetGluArgGlnArgLeuIleGlnSerValProAspSerValThrProLeu 420
DB 1260 GATGATGTATGAAAGGACAGACTCATTTAGTCAGTGCCTGATTCGTGACGCCACTA 1319
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
DB 1320 ATTATTCATGAAGAAACAACAGACATCTGATTAATATCATGACATCTTTCATGTTT 1379
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
DB 1380 CCCCAAGTCAAGAAAGAAATGAGTTTATTGCTCTGATGCAAAACAGGTTTC 1439
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArgSerSerGly 480

DB 1440 CGTCATTTATACAAATTAATCACTATTTTAAAGAAACAAATTAACGATCCAGTGT 1499
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB 1509 GGGCTGCTGCTCCCAAGTATTCAGAGTCTCTATCAAGAGAGATAGCAATTACAGT 1559
QY 501 GlyLutProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB 1560 GGTAAATGGAAAGTTCTTGGCGGCGATGGATCTAATATCCAAAGTTGATGAATCGAAAG 1619
QY 521 LeuValTyrPheGlnGlyThrLysAspSerProLeuGlnIleHisLeuTyrValIleSer 540
DB 1620 CTGGTATATTTTGAAGGACCAAGACTCCCTTTAGAGCTACCTGTAGCTAGTCACT 1679
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB 1680 TACGTAATCTCGAGAGAGTGAACAAGGTGACTGACCGTGGCTACTCATCTTGCTGTC 1739
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
DB 1740 ATCAGTCAAGCACTGTGACTTCTTATTAAGTAAGTAAGTAACCAAGAAATCCACACTGT 1799
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
DB 1800 GTGTCCCTTACAAAGCTATCAAGTCTGAAAGATGACCCAACTTGCAAAACAAAGAAATTT 1859
QY 601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620
DB 1860 TGGGCAACCATTTTGATTTGACAGAGTCTCTTCTGCTACTATACCTCCACGAAATTTTC 1919
QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
DB 1920 TCTTTTGAAGTACTACTGATTAATTAATGGAAGGCTCTCAACCTCATATATCTTA 1979
QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
DB 1980 CAGCTGGAAAGAAATATCTTACTGTGCTGTTCATATATGATGATGCTCTCAG----- 2030
QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
DB 2030 ----- 2030
QY 681 TyrValValValValIleAspAsnArgGlySerCysHisAlaGlyLeuLysPheGluGly 700
DB 2030 ----- 2030
QY 701 AlaPheLysTyrLysMetGlyGlnIleGluThrLeuAspArgIleValGluGlyLeuGlnTyr 720
DB 2030 ----- 2030
QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleGlyTyrPserTyr 740
DB 2030 ----- 2030
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2031 -----GTTCCTATT 2039
QY 761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
DB 2040 GCTGGGGCCCAAGTCACTCTGAGATCTTCTATGATACAGGATACACCGAAAGCTTATAG 2099
QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
DB 2100 GGTCAACCTGACCAAGATGAAACAGGGCTATTAAGTCTGTGGCCATCGAACACAA 2159
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
DB 2160 AAGTTCCCTCTGAAACCAATCGTTTACTGCTTAACTAGGTTTCTGGATGAGAAATGTC 2219
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyAlaGlyLysProTyrAsp 840

OY	331	AAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle	340
Db	1174	GCMAATCCCTAAAGTCACTTTTAAGATGTCAAGAAATTAATGATTGAATGCTGAAGAAAGATC-	1233
OY	341	IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu	360
Db	1232	-----	1233
OY	361	TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp	380
Db	1232	-----	1233
OY	381	ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu	400
Db	1232	-----	1233
OY	401	AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu	420
Db	1232	-----	1233
OY	421	IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe	440
Db	1232	-----	1233
OY	461	ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly	480
Db	1232	-----	1233
OY	481	GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer	500
Db	1232	-----	1233
OY	501	GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg	520
Db	1233	-----CAAGTGATGAAGTCAAGAG	1255
OY	521	LeuValTyrTrpPheGluGluYThrLysAspSerProLeuGlnHisHisLeuTyrValValSer	540
Db	1254	CTGGATATTTTGAAGGACACCAAAACCTCCCTTTAAGACATCACTGTACGTACTACTGCT	1313
OY	541	TyrValAsnProGlyGlyValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys	560
Db	1314	TACGTAAATCCTCGAGAGGTGACAAAGCTGACCTGCTGCTACATTTCTTGCTTC	1373
OY	561	IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys	580
Db	1374	ATCAGTCAAGCACTGTGACTTCTTTTAAAGTAAGTAACTAAACAGAAAGATTCACACTGT	1433
OY	581	ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe	600
Db	1434	GTGTCTCCCTTAACAGCTATCAAGTCTTAAGATGATGACCCAACTTGCAAAACAAAGGAATTT	1493
OY	601	TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe	620
Db	1494	TGGGCGACACATTTTGATTGATTCAGACAGTCTCTTCTCTGACTATATCTCCCGAAGAAATTTTC	1553
OY	621	SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu	640
Db	1554	TCTTTTGAAGTACTACTGATGATTAACATTTGAAGGAAGCTCTCAACAGCTCATGATCTGA	1613
OY	641	GlnProGlyLysLysTyrProThrAlaLeuPheIleTyrGlyGly--ProGlnValGlnL	660
Db	1614	CAGCTGGAAAGAAATATCTTAAGTGTCTGTTCATATATGATGATGCTCTCAAGTGCAGGT	1673
OY	660	euvAlaAsnAsnArgPheLysGlyValLysTyrPheArgLysAsnThrLeuAlaSerLeuG	680
Db	1674	TGGTGAATAATTCGGTTTAAAGAGTCAAGTATTTCCGGCTTGAATTAACCTTACCTCTCTCAG	1733

[illegible]

Search completed: October 15, 2003, 17:56:50
Job time : 780.054 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:29:02 ; Search time 138.181 Seconds
 (without alignments)
 2817.309 Million cell updates/sec

Title: US-10-070-464-1
 Perfect score: 4700
 Sequence: 1 MAAAMETEQLGYEIFETADC.....HLHLVQENIGSRIALAKVI 882

Scoring table:	
BLOSUM62	
Xgapop 10.0 ,	Xgapext 0.5
Ygapop 10.0 ,	Ygapext 0.5
Fgapop 6.0 ,	Fgapext 7.0
Delop 6.0 ,	Delext 7.0

Searched: 569978 seqs, 220691566 residues
Total number of hits satisfying chosen parameters: 1139956

```

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

Command line parameters:
-MODEL=ftmame+ p2n.model -DEV=xlh
-D=/cgn2.1/USPPO.spool/US10070464/runat.15102003.113554.24845/app_query.fasta.1.26522
-QB=Issued Patents NA -GPMT=fastacp -SUFFIX=xmi -MINMATCH=0.1 -LOOCL=0
-LOOPEXT=0 -UNITS=515c -STAR=1 -END=1 -MATRIX=bls10m62 -PRANS=humand40.cdi
-LIST=45 -DOCLINK=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=PCO -NORF=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464.GCGN.1.1.146/runat.15102003.113554.24845 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESCORE -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-TIMEOUT=120 -WARN TIMEOUT=0 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELTOP=6 -DELXT=7

```

```
Database : Issued Patents NA:
1 : /cgn2.6/prodata/1/lna/5A.COMB.seq:*
2 : /cgn2.6/prodata/1/lna/5B.COMB.seq:*
3 : /cgn2.6/prodata/1/lna/6A.COMB.seq:*
4 : /cgn2.6/prodata/1/lna/6B.COMB.seq:*
5 : /cgn2.6/prodata/1/lna/PCUTS.COMB.seq:*
6 : /cgn2.6/prodata/1/lna/bacillus.seq:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	755.5	16.1	823	US-09-280-116-171	Sequence 171, Appl1
2	597.5	12.7	5496	US-09-462-284-1	Sequence 1, Appl1
3	529	11.3	3407	US-10-002-593-5	Sequence 5, Appl1
4	525	11.2	2924	PCT-US93-07923-1	Sequence 1, Appl1
5	454	9.7	2815	US-08-230-919A-1	Sequence 1, Appl1
6	454	9.7	2815	US-08-619-280A-1	Sequence 1, Appl1
7	454	9.7	2815	US-08-960-391-1	Sequence 1, Appl1
8	404	8.6	612	US-09-352-184-31	Sequence 31, Appl1
9	339.5	7.2	4982	US-08-639-103B-1	Sequence 1, Appl1
10	339.5	7.2	4982	US-09-229-059-1	Sequence 1, Appl1
11	339.5	7.2	4982	US-09-628-133-1	Sequence 1, Appl1
12	328.5	7.0	657	US-09-221-017B-646	Sequence 646, Appl1

C	13	328	7.0	543	4	US-09-221-017B-253	Sequence 251, App
C	14	282	6.0	1896	3	US-09-016-080-2	Sequence 2, Appl
C	15	278.5	5.9	535	4	US-09-389-681-428	Sequence 428, App
C	16	278.5	5.9	535	4	US-09-620-856B-428	Sequence 428, App
C	17	278.5	5.9	535	4	US-09-433-826B-428	Sequence 428, App
C	18	278.5	5.9	535	4	US-09-604-287A4-1010	Sequence 428, App
C	19	232	4.9	3085	4	US-09-421-017B-1010	Sequence 1010, App
C	20	193	4.1	502	4	US-09-280-116-172	Sequence 172, App
C	21	182	3.9	1869	2	US-08-664-666A-1	Sequence 1, Appl
C	22	182	3.9	1869	2	US-09-066-285-1	Sequence 1, Appl
C	23	182	3.9	1869	3	US-09-261-006-1	Sequence 1, Appl
C	24	182	3.9	1869	3	US-08-951-088-1	Sequence 1, Appl
C	25	182	3.9	1869	4	US-09-609-566-1	Sequence 1, Appl
C	26	182	3.9	1869	4	US-09-609-570-1	Sequence 1, Appl
C	27	182	3.9	1869	4	US-09-427-372-1	Sequence 1, Appl
C	28	182	3.9	1869	4	US-09-693-552-1	Sequence 1, Appl
C	29	175.5	3.7	1845	4	US-09-253-991A-3485	Sequence 3489, App
C	30	157.5	3.8	1032	4	US-09-221-017B-962	Sequence 962, App
C	31	133	2.4	815	4	US-09-724-623-17	Sequence 17, Appl
C	32	133	2.8	1230025	4	US-09-198-057B-58	Sequence 1, Appl
C	33	128.5	2.7	904	4	US-09-221-017B-58	Sequence 58, Appl
C	34	128	2.7	536165	4	US-09-214-808-1	Sequence 1, Appl
C	35	125.5	2.7	5837	1	US-07-666-340-1	Sequence 1, Appl
C	36	125.5	2.7	5837	1	US-08-004-139B-1	Sequence 1, Appl
C	37	125.5	2.7	5837	1	US-08-117-491-1	Sequence 1, Appl
C	38	125.5	2.7	5837	1	US-08-271-364A-1	Sequence 1, Appl
C	39	125.5	2.7	5837	2	US-08-811-492-1	Sequence 1, Appl
C	40	125.5	2.7	5837	1	US-08-222-715B-1	Sequence 1, Appl
C	41	125.5	2.7	5837	5	PCIT-US96-10545A-1	Sequence 1, Appl
C	42	124	2.6	2636	1	US-08-127-689-1	Sequence 1, Appl
C	43	124	2.6	2636	3	US-09-366-169-1	Sequence 1, Appl
C	44	118.5	2.5	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C	45	117.5	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appl

ALIGNMENTS

```

1 RESULT 1
2 US-09-280-116-171
3 / Sequence 171, Application US/09280116A
4 / Patent No. 631427
5 / GENERAL INFORMATION:
6 / APPLICANT: Robison, Keith E.
7 / TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
8 / FILE REFERENCE: 5800-24, 035800/176965
9 / CURRENT APPLICATION NUMBER: US/09/280,116A
10 / CURRENT FILING DATE: 1999-03-26
11 / NUMBER OF SEQ ID NOS: 268
12 / SOFTWARE: PatentIn Ver. 2.0
13 / SEQ ID NO 171
14 / LENGTH: 923

```

Alignment Scores:		
Pred. No.:	1,96e-77	Length: 822
Score:	755.50	Matches: 16
Percent Similarity:	88.95%	Conservative: 5
Best Local Similarity:	86.32%	Mismatches: 5
Query Match:	16.07%	Indels: 16
DB:	4	Gaps: 2

Db 63 CATTGACTTAGATCGTGGGATCCACGGCTGCTCTATGAGATACCTCTCCCTGAT 122
Qy 747 TAlaIeUmetGlnArgSerAspIlePheArgValAlaIleAlaGlyValaProValThre 767
Db 123 GGCATTAAAGCAGAGGTGAGATATCTTCAGGGTTCCTATGCTGGGTCCCACTCATCTT 182
Qy 767 UTPrIlePheTyArgPThrGlyTyArgThrGluArgTyMetGlyHisProAspGlnAsnG 787
Db 188 GTGGATCTTCTATGATACAGAGATACAGGAACTTATATGGGTCACTGACCAAGATGA 242
Qy 787 uGlnGlyTyTyLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAs 807
Db 243 ACAGGGCTATATCTTAGATCTGTGGCCATGACGAGAAAGTTCCCTGGAACCAAA 302
Qy 807 nArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerI 827
Db 303 TCGATTACGCTCTTACAGAGGTTCTCGATGAGAAATGTCATTTTGGACATACAGATAT 362
Qy 827 eLeuLeuSerPheLeuValArgAlaGlyLysProTyArgPLeuGln---IleTy----- 844
Db 363 ATTACTGAGTCTTTTAGTAGAGGGCTGGAAGCAGATATGATTTACAGTATCTTATTTT 422
Qy 845 -----ProGlnGluArgHis-SerIleArgVal-ProGluSer-GI 857
Db 423 GTTGTGTGATAGATCTATCTCAGGAGAGACACCGGCATTAAGAGTTCCCTGATCGGG 482
Qy 857 uGlnHisTyArgIleLeuHisLeuLeu-HisTyArgGlnGlu-AsnLeuGlySer-ArgI 876
Db 483 ABAACCTTATGAACTGCATCTTTTGGACCTACCTTCAAGAAAACCTTGGATCACTTAT 542
Qy 876 eAlaAlaLeuLysVal 881
Db 543 TGCTGCTCTTAAATT 558

RESULT 2
US-09-462-284-1
; Sequence 1, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Aifolter, Michael
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; TITLE OF INVENTION: ASPERGILLUS ORYZAE
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Fungus
US-09-462-284-1

Alignment Scores:
Pred. No.: 9.56e-58 Length: 5496
Score: 597.50 Matches: 208
Percent Similarity: 42.67% Conservative: 112
Best Local Similarity: 27.73% Mismatches: 152
Query Match: 12.71% Indels: 152
DB: 4 Gaps: 33

US-10-070-464-1 (1-882) x US-09-462-284-1 (1-5496)
Qy 145 IlegIyThrValGlyIleAlaSerTyArgPThrHisGlnGlySerGlyThrPhe----- 162
Db 2239 GTCCGTGCTGCTGGCGTCAACCA-----CACCAAGCAGTATCGGCATCTCTTCT 2289
Qy 163 -----LeuPheGlnAlaGlySerGlyIleTyThrHisValLysAspGlyProGlnGly 180

Db 2290 TCCGATTTATTCGTCAGAGTGTGAGTCACTCAAGTCCG----- 2331
Qy 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 2332 -----GCCCTGATGCCGATCA 2349
Qy 201 AspPro-----LysLeuCysProAlaAspPro-----AspTrp-IleAlaPheIleHis 216
Db 2350 GGAAGGTGATATTCAATATGACCAATGAGCCCGCGCAATGACATCGCTTTGTGTTG 2409
Qy 216 sSerAsnAsp-----IleTrpIleSerAsnIleValThrArgGluLysArgLeuThr 234
Db 2410 CGAGATATGACCTTATATGCTGGGATATGATGCTTACTGCC----- 2452
Qy 234 rTyValHisAsnGluLeuAlaAsnMetGluLysAla-----ArgSerAlaGI 251
Db 2453 -----ATTACGATATGATGATGCGCCGACATGTCACCG 2487
Qy 251 yValAlaThrPheValLeuGlnGluGluPhe-----AspArgTySerGlyTyTrpTr 269
Db 2488 CGTGCCGAGCTGGATCTATGAAAGAGAGATCTCGGCATCGCTACGCG--TTGGTGT 2544
Qy 269 pCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyArgL 289
Db 2545 CTCGCCAGATGGTGA-----TATCTGGCTTACTTGAG 2577
Qy 289 uAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThrArgAr 309
Db 2578 CTTCAATAGACATGGGGTTCCGACCTACACCGTTCAATATATGATTAACCAAGAGAT 2637
Qy 309 gAlaAspSerPhe-----ArgTyProLysThrGlyThrAlaAspR 323
Db 2638 CCGCTCCGGATATCCATGAGAGCTGAAGATTAAGATCCCAAGGTGTCGACAGAGATCC 2697
Qy 323 oLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleAspVa 343
Db 2698 GACCGTGAAGTTGAT-----CTGCTTAACAT 2724
Qy 343 ILeAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGluTyTrIleAl 363
Db 2725 CGCTAGCAAGAGGAGGAGAGAGCCGACATGACCGCTTCGATCAAGCATGATGATCAT 2784
Qy 363 AArgAlaGlyTrpThrProGluGlyLysTyArgAlaTrpSerIleLeuLeuAspArgSerGI 383
Db 2785 T-----GGGAGGTTGGCTGG-----CTCACTGATATCTCACAC 2817
Qy 383 nThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGluAspAspVa 403
Db 2818 CACC-----GTTGCTGCTAAGCGCTTCAACCGTGTCCAGAC----- 2854
Qy 403 IMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeuIleIleTy 423
Db 2855 -CAGCAAAAGTCGTGCGGTGATGATCTCGCAACCAAGCTACT---GTCATACAGCA 2910
Qy 423 rGlnGluThrThrAspIleTrpIle---AsnIleHisAspIlePheHisValaPheProGI 442
Db 2911 CCGAGATGAGACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2970
Qy 442 nSerHisGlnGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPheArgHis 462
Db 2971 CAAGCGTCGCAAGAGATGCTCTACTATGACATCTCTGACCTTGGAGTGGCGCA 3030
Qy 462 eLeuTyTrIleLysIleThrSerIleLeuLysGluSerLysTyTrIleAspArgSerGlyLe 482
Db 3031 TCTGTATCTCTTCCC-----GTTTGGGGGGGCGCA 3060
Qy 482 uProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSerGlyGI 502
Db 3061 ACCTATCCCA-----CTTAACCAAGGCGCA 3084
Qy 502 uTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuVa 522


```

Db      3085 CTGGAGGCTCAG-----TCTATTCTGAGTATTGATCAGAAAGCCAGTTGGT 3132
QY      522 1TyrPheGluGluYThrLysAspSerProLeuGluHisLeuYValValSerTyrVa 542
Db      3133 GTACTACTGTCGTGATCAGACACAGCAGCCAGAGCCATCTCTACTCTCTCTAT-- 3190
QY      542 1AsnProGluGluValThrArgLeuThrAspArgGlyTyrSer-----HisSerCysCy 560
Db      3191 -TCCACGTTGGGGTCCGCCCTGTCGTGACAGACACCGTTGGCGGTACTGTCGTTCC 3249
QY      560 s1LeSerGlnHisCysAspPhePhe1LeSerYtyrSerAsnGlnLysAsnProHisCy 580
Db      3250 CTTCTCCCGGAACCTGGGCTACTACATCTTCACATACGAGAGCCAGACGTACCTAC- 3307
QY      580 sValSerLeuYtyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPh 600
Db      3308 -CAGGAACCTTACAGACCAACAGT-----ACCAAAACCACT 3342
QY      600 eThrAlaThrL1LeuAspSerAlaGlyProLeu-----ProAspTyrThrProPr 617
Db      3343 C---CGCACATCAGCGACAGCCAAAGTACTCGAGCAATCAAGAGCTATGCTTGC 3399
QY      617 oGlu1LePheSerPheGlu-----SerThrThrGlyPheThrLeuYtyrGlyMetLeuTy 635
Db      3400 CAACATCTACCTACTTGAAGCTTCCCTCCCTCCGAGAAACCTCAATGTATGATCAGCG 3459
QY      635 rLysProHisAspLeuGlnProGlyLysLysYtyrProThrValLeuPhe1LeTyrGlyG 655
Db      3460 CTTACCCCCCGGGTCTCTCCCGGATTAAGAGTACCCATATCTTTCACCCCATAGCGCG 3519
QY      655 yProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysYtyr---PheArgLeuAs 674
Db      3520 CCCAGCGCCCGCAAGAGTACCAAGAGATGGCAGCGCTCAATTTCAAGGCTATGATGCG 3579
QY      674 nThrLeuAspSerLeuGlyYtyrValValVal1LeAspAsnArgGlySerCysHisArg 694
Db      3580 CTCACAGCGCAACTCGAGTACGTAACTGACGTGCAACCGCGGACAGGTTTCAA 3639
QY      694 gGlyLeuLysPheGluGluYAlaPheLysYtyrLysMetGlyGln1LeGlu1LeAspAsp 714
Db      3640 AGGAGCAAGTCCCTCCGCTCGGCTACCGCGGCAACCTCGGCTCTCGAAGCAAGACCA 3699
QY      714 nValGluGlyLeuGlnYtyrLeuValAspArgYtyrAspPhe1LeAspLeuAspArgVal 734
Db      3700 GATCTACCGCGCGCAACAG---GCGGCCAATCCCTCGATCGATGACAGACCATCGG 3756
QY      734 y1LeHisGlyYtyrPserYtyrGlyYtyrLysSerLeuMetAlaLeuMetGlnArgSerAs 754
Db      3757 CATCTGGGGCTGAGTTCGGAGGCTACTTTCACAGCAAGCTCTCGAAGACAGACGCGG 3816
QY      754 r1LePheArgValAla1LeAlaGlyAlaProVal1ThrLeuTyr1LePheYtyrAspThrG 774
Db      3817 TGCCTTCAATTAGAGATCATACCGCCCTGTCTTCTGACTGGCGTTCTTACAGTCAAT 3876
QY      774 yTyrThrGluArgYtyrMetGlyHisProAspGlnAsnGluGlnGlyYtyrYtyrLeuGlySe 794
Db      3877 GTACCGGAGGCTCATGATAAGACCCCTCGACCAATGAGAGAGGCTAGACAGACGCGC 3936
QY      794 rValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGly 814
Db      3937 CGTCT---CGCAGAGCTGACGGGTTCAAGAACTCAGAGGCGGATTTCTTGATTCACAGCG 3993
QY      814 yPheLeuAspGlnAsnValHisPheAlaHisThrSer1LeLeuLeuSerPheLeuValAr 834
Db      3994 AACGGGCAACATACATTCATTTCCAGAACTCGGCTGGCTGGTGATCTCTTATGGG 4053
QY      834 gAlaGlyLysProYtyrAspLeuGln1LeTyrProGlnGluArgHisSer1LeArgValPr 854
Db      4054 CGATGGC-----GTCTCTCTGAGAAAGTCATTCGCAATGGTTTAC 4095
QY      854 oGluSerGlyLysIleYtyrGluLeuHis 863
Db      4096 AGACTCAGACCAAGCAATCACTTACAT 4123

```

```

RESULT 3
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOGENSIN CONVERTI
; FILE REFERENCE: Acty Docet No. 6586198 1242/48/2
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US/10/002,593
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.: 3,44e-50 Length: 3407
Score: 529.00 Matches: 186
Percent Similarity: 41.96% Conservative: 109
Best Local Similarity: 26.46% Mismatches: 280
Query Match: 11.26% Indels: 128
Gaps: 32
DB: 4

US-10-070-464-1 (1-882) x US-10-002-593-5 (1-3407)

QY      206 Pro1aAspProAspTyr1Le-----AlaPhe1LeHis 216
Db      520 CCMAACACACACAGTGGGTGCATGTCACCAAGTGGTCAATTAATGGCATATGTTGG 579
QY      217 SerAsnAsp1LeTyr1LeSerAsn1LeVal1ThrArgGluGluArgLeuThrYtyrVal 236
Db      580 AACCAATGACATTTATGTTAAATTAAGCAAAATTTACCAAGTTTACCAATTCACATGGAGC 639
QY      237 HisAsnGluLeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheVal 256
Db      640 GGG-----AAAGAGATTAATTAATTAATGAGATACTACGCGTT 681
QY      257 LeuGlnGluGlu---PheAspArgYtyrSerGlyYtyrTyrProLysProLysAlaGluThr 275
Db      682 TATGAAAGAGAAAGTCTGACGTGCTTCTGCTGTGTGTGTGTCTTCAACGCGCACTTT 741
QY      276 ThrProSerGlyGlyLys1LeuArg1LeuLysYtyrGluGluAsnAspGluSerGlyVal 295
Db      742 -----TTAGCATATGCCCAATTTTAAACGACACAGAAAGTC 774
QY      296 Glu1Le1LeHis-----ValThrSerProMetLeuGluThrArgArgAlaAspSer 312
Db      775 CCACCTATTGAATACTCTTCTACTCTGATAGTACTGCGAGTACCGCAAGCTGACG 834
QY      313 PheArgYtyrProLysThrGlyThrAlaAsnProLysVal1ThrPheLysMetSerGluLe 332
Db      835 GTTCCATATCCAAAGGAGAGCTGTGATCAATCTGTAAGATTCTTTGTT-----GTA 888
QY      333 Met1LeAspAlaGluGlyArg1Le1LeAspVal1LeAspLysGluLeu1LeGlnProPhe 352
Db      889 AATACAGACTCTTCACTCACTCAATGCAATCTTCATCAATATCACTGCTCTGCT 948
QY      353 Glu1LeuPheGluGlyValGluYtyr1LeAlaArgAlaGlyTyrThrProGluGlyLys 372
Db      949 TCTATGTTGATA---GGGAGTACACTCTTGTGTATGATGACATGGCA----- 993
QY      373 TyrAlaTyrSer1LeLeuLeuAspArgSerGlnThrArgLeuGln1LeValLeu1LeSer 392
Db      994 -----ACACAGAAAGAAATTTCTTTCAGTGGCTC--- 1023

```


STRANDEDNESS: double		
TOPOLOGY: linear		
PCT-US93-07923-1		
Alignment Scores:		
Pred. No.:	7,666-50	Length: 2924
Score:	525.00	Matches: 186
Percent Similarity:	41.58%	Conservative: 108
Best Local Similarity:	26.31%	Mismatches: 277
Query Match:	11.17%	Indels: 136
DB:	5	Gaps: 33
US-10-070-464-1 (1-882) x PCT-US93-07923-1 (1-2924)		
Qy	206	ProAlaAspProAspTrrPle-----AlaPheIleHis 216
Db	455	CCAAACAAACACACAGTGGGTGCATGTCACCGATGGGTCAATATGGCATATGTTGG 514
Qy	217	SerAnaPpIleTrrPleSerAnaIleValThraTgGluGluAraArgLeuThryVal 236
Db	515	AACAATGACATTATGATTAAATTTGAACCAAAATTTACAGATTACAGATACACAGACG 574
Qy	237	HisAnaGluLeuAlaAsmGluGluAspAlaArgSerAlaGluValAlaThrPheVal 256
Db	575	GGG-----AAAGAGATATATATATATATATGATATACATGACTGGTT 616
Qy	257	LeuGluGluGlu---PheAspArgTyrSerGlyTyrTrrPcyAspPolsAlaGluThr 275
Db	617	TATGACAGAGAAAGTCTTCAGTCCCTACTGCTGCTGGTGTCTCCAAACGCACATTTT 676
Qy	276	ThrProSerGlyGlyLysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluVal 295
Db	677	-----TTAGCATATGCCCAATTTAAACACACACAAAGTTC 709
Qy	296	GluIleIleHis-----ValThrSerProMetLeuGluThrArgArgAlaAspSer 312
Db	710	CCACTTATTTGAATTACTCCTTCTACTCTGTATGATGATCAGTCCAGTACCCAAAGCTTACGG 769
Qy	313	PheArgTyrProLysThrGlyThraAlaAspPolsValThrPheLysMetSerGluIle 332
Db	770	GTTCCATATCCAAAGCAGCAGCTGTGATTCACACGTGAAGTTCTTTGTT-----GTA 823
Qy	333	MetIleAspAlaGluGlyAraGlyIleLeaAspValIleAspLysGluLeuIleGlnProPhe 352
Db	824	AATACAGACTCTCAGCTCAGTCAACCAATGCAACTTCATCAATCACTGCTCTGCT 883
Qy	353	GluIleLeuPheGluGluValGluTyrIleAlaArgAlaGlyTrrPthProGluGlyLys 372
Db	884	TCTATGTTGATA---GGGGATCACACTTGTGTGATGTGCATCGGACA----- 928
Qy	373	TyrAlaTrrSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSer 392
Db	929	-----ACACAGAAAGAAATTTCTTTGCACGCGCTC--- 958
Qy	393	ProGluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSer 412
Db	959	-----ACGAGAGATTACAGAAC 973
Qy	413	ValProAspSerValThrProLeuIleIleTyrGluGluGluThrThrAspIleTrrPheLsn 432
Db	974	TAT-----TCGGTCATGATATTTGTGATCTATGATGATGATCAGTGAAGATGG---AAC 1024
Qy	433	IleHisAspIlePheHisValPheProGlnSerHisGluGluGluIleGluPheIlePhe 452
Db	1025	TGC-----TTAGTGGCAGCAGCAACACATTGAATGATGATCTAGTGGCTGG 1066
Qy	453	AlaSerGluCysLysThrGlyPheArgHis-----LeuTrrLys 465
Db	1070	GTTGGAGATTTAGGCTTCAGAACCTCATTTTACCTTGATGGTAATAGCTTCTACAAG 1128
Qy	466	IleThrSerIleLeuLysGluSerLysTrrLysAspSerSerGlyGlyLeuProAlaPro 485
Db	1130	ATCATCATCC-----AATGAAGAAAGTTACACACATTT----- 1162

QY	486	SerSepPheIysCysProIleuGluIleAlaIethrSerGlyIuTrpGluVal	505
Db	1163	TGCTATTTTCCAAATAGATMAAAAGACGCACATTTATTACAAAAGGACCTGGGAAGTTC	1222
QY	506	LeuGlyATrGhiIseGlySerAsnIleGluValAspGluValArgGluValTyrrheGlu	525
Db	1223	ATCGGG-----ATAGAACTTAAACCAAGTATCTTACTACTACTTACT	126
QY	526	GlyThr---LysAspSerProLeuGluIuIshIleuTyrrValIalSerTyrrValAsnPro	544
Db	1268	AATGAATATTAAGAATGCCAGAGGAAGAAATCTTTATAAATC-----	1311
QY	545	GlyIuValThrArgLeuThrAspArgGlyTyrrSerHisSerCysCysIleSer-----	562
Db	1313	-----CAACTTATGAC-----TATACAAAGTACATGCTCTCAGTTGTGAG	135
QY	563	-----GlnHisCysAspPhePheIleSerLysTyrrSerAsnGluLysAsnProHis	579
Db	1355	CTGAATCCGGAAGGTGTCACTACTTCTGTGTATTCATTCAGTAAAGAGCGCAAGTATTT	1411
QY	580	-----Cys-----ValserLeuTyrrLysLeuSerSerProGluAsp	592
Db	1415	CAGCTGAGATGTTCCGCTCTGCTGCTGCTCTCTACTTACACAGACAGCGTGAATGAT	1474
QY	593	ProthrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro	612
Db	1475	AAAGCGCTAGAGATCTCGAAAGAC--AATTCAGCTTTTGATTAATAAG--CTGAG	1522
QY	613	AspTyrrThrProGlu-----IlePheSerPheGluSerThrThyGlyPheThr	629
Db	1528	AATGTCAGATGCCCTCCAAAACTGACCTTATTTGTAAGAAACAATAATTT---158	158
QY	630	LeuTyrrGlyMetLeuTyrrLysProHisAspLeuGluProGlyLysLysTyrrProThrVal	649
Db	1583	TGCTATCAGATGATCTTCCTCTCAT---TTTGATTAATCCAGAAATATCTCTACTA	1633
QY	650	LeuPheIleTyrrGlyGlyProGluValGlnLeuValAsnAsnArgPheLysGlyValLys	669
Db	1640	TTAGATGTGTATGACAGGCCCATGTAGTAA-----AAAGCAGACACT	1681
QY	670	TyrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTyrrValValValVal---	685
Db	1682	GTCTTCACACTGMACTGGGCCACTTACTCTTCACACACAGAAACATTTACTACTAGC	1744
QY	686	IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaIlePheLysTyrrLys	705
Db	1742	TTTGATGCGCAGAGGAAGTGTTCACAAAGATAAATCATGCATGCAATCAACAGAGA	1801
QY	706	MetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrrLeuAlaSerArgTyrr	725
Db	1802	CTGGGACATTTGAAGTTGAAGATCAAAATTAAAGCAGCCAGCAAAATTT--TCAAAAATG	1856
QY	726	AspPheIleAspLeuAspArgValGlyLysIleHisGlyTrpSerTyrrGlyGlyTyrrLeuSer	745
Db	1859	GGATTTTGACAAACAACGAATTCGCAATTTGGGCTGTCTATATGACGGTACTTAAC	1911
QY	746	LeuMetAlaLeuMetGlnArgSerThrAspIlePheArgValAlaIleAlaGlyAlaProVal	765
Db	1919	TCAATGTGCTCCGTGATCAGAAAGTGGCGCTGTTCAGGTGGAATGACCGTGGCGCTCTTA	1978
QY	766	ThrLeuTrpIlePheTyrrAspThrGlyGlyTyrrThrGluArgTyrrMetGly-----HisPro	783
Db	1979	TCCGGTGGAGTACTATAGCTACTGTGACACAGACCTTACATAGGGCTCCCAACTCA	2038
QY	784	AspGlnAsnGluGlnGlyTyrrTyrrLeuGlySerValAlaMetGlnAlaGlyLysPhePro	803
Db	2039	GAAGCAACCTTGACCATTAAGAAATTCACACAGTCATGAGCAGACGTGAATAATTT---	2095
QY	804	SerGluProAsnArgLeuLeuLeuLeuLysGlyPheLeuAspGluAsnValHisPheAla	823
Db	2096	---AAACAAGTTGAGTACTCTTATTCATGGAACAGACAGATGATTAACGTTTCACTTTCAG	2152

QY 824 HistHsrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrPheLeuGlnIle 843
Db 2153 CAGTAGCTCAGATGCTCCAAAGCCCTCGATGTTGGATGTCATTTCCAGCAATGTCG 2212
QY 844 TyrProGlnIuArgHisSerIleArgValProGlnSerGlyLuhHisTyrGluLeuHis 863
Db 2213 TATACTGATGAAGACCATGATAGTACGACACACGACACACATATATATACCCAC 2272
QY 864 LeuLeuHisTyrLeuGlnIu 870
Db 2273 ATGACGCACTTCATAAACAA 2293
RESULT 5
US-08-230-491A-1
Sequence 1, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.;
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FELPE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-230-491A-1
Alignment Scores:
Pred. No.: 1.15e-41 Length: 2815
Score: 454.00 Matches: 177
Percent Similarity: 41.92% Conservative: 111
Best Local Similarity: 25.76% Mismatches: 278
Query Match: 9.66% Indels: 122
DB: 1 Gaps: 27
US-10-070-464-1 (1-882) x US-08-230-491A-1 (1-2815)
QY 212 IleAlaPheIleHisSerAsnSerPheIleTyrPheSerAsnIleValThrArgGluArg 231
Db 692 TTAGATATAGTGTATCAAAACATATCTATTGAAACAAAGACGAGATCCACCTTTT 751
QY 232 ArgLeuThrTyrValHisAsnGluLeuAlaAsnMetGluIuAspAlaArgSerAlaGly 251
Db 752 CAATAAACAATT-----AATGAAAGAAATAATTAATTTATATGGA 793
QY 252 ValAlaThrPheValLeuGlnIuGluPhe-----AspArgTyrSerGlyTyrTrpTyr 269

Db 794 ATCCAGACTGGGTTTATGAAGAGAAATGCTTCTACAAATATGCT---CTCTGGTG 850
QY 270 CysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGluGln 289
Db 851 TCTCCT-----AATGAAATTTTG-----GCATATGCGGAA 883
QY 290 AsnArgGluSerGluValGluIleIleHisValIleThrSerProMetLeuGlnThr---Arg 308
Db 884 TTTAATGATTAAGATATACAGATATGCTTATGCTTATATGAGGATACAAATATCCT 943
QY 309 ArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLys 328
Db 944 AGAACATTAATATTCATACCCAAAGCTGAGCTAAGAAATCCCTTTGGGATA--- 1000
QY 329 MetSerGluIleMetIleAspAlaGluArgIleIleAspValIleAspLysGluLeu 348
Db 1001 -----TTATATATGATATACCACTTACCTGGCTAT 1030
QY 349 ILeGlnProPheGluIleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTyrThr 368
Db 1031 GTAGTCCCCAGAAAGTGGCTGTTCCAGCAATG-----ATAGCCTCAAGTATAT--- 1081
QY 369 ProGlnGlyLysTyrAlaThrPheSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIle 388
Db 1082 -----TATTCAGTTGGCTACAGCGGCTTACTGATGAGACGATAGTTGGACGTG 1132
QY 389 ValLeuIleSerProGluLeuPheIleProValGluAspAspValMetGluArgGlnArg 408
Db 1133 CTAAAAAG-AGTCCAGAAATGTTCCGCTCCTGC-----TATATGAGACTTCAGGGA 1182
QY 409 LeuIleGluSerValProAspSerValThrProLeuIleIleTyr-GluGlnThrThrs 428
Db 1183 AGACTGGCAGACATGGATGTTCCAAAGACCA---CGAGCATATGAAGAAAGCAAGAC 1239
QY 428 PLeuTrpIleAsnIleHisAspIlePheHisValPheProGlnSerHisGluGluGluI 448
Db 1240 TGGATGGGCTGCT-----GATTTCTTTGTTTCAGAGCAGATTTCCAGTATAGCCAT 1293
QY 448 eGluPheIlePheAlaSerGluCysLysThrGlyPheArgHisLysLeuTyrLysIleThse 468
Db 1294 TTCGTAACTAACAAATATTAGTACAGAGATGAGTCAAAACATATTCACATATATC----- 1348
QY 468 rIleLeuLysGluSerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPh 488
Db 1348 ----- 1348
QY 488 eLysCysProIleLysGluGluIleAlaIleThrSerGlyGluTyrGluValLeuGluYar 508
Db 1349 -AAAGACACTGGAAATGCTATTCAAATTCAGAGTGGCAGTGGAGGCCATA----- 1402
QY 508 gHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPheGluGlyThry 528
Db 1403 -----AATATA-----TTCAAGATAACAA 1422
QY 528 sAspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsnProGlyGluValh 548
Db 1423 GGAATCACTGTTTATTTCTAGCAATGAATTTGAAATACCTCGAAGAAACATCTTA 1482
QY 548 rArgLeuThrAspArgGlyTyrSerHisSer---CysCysIleSerGlnHis----- 564
Db 1483 CAGATTTAGCATTTGAGAGCTATCTCCAAAGCAAGAGTGTACTGTGCATCTAAGAA 1542
QY 565 -----CysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSe 582
Db 1543 AGAAAGGCGCCAAATTTACACAGCAAGTTTCAGCAGCTACCGCAAGTACTAGCACTTGT 1602
QY 582 rLeuTyr-----LysLeuSerSerProGluAspAspProThrCysLysThry 598
Db 1603 CTGCTACGCGCCAGGATCCCATTTCCACCTTATATATGAGCAGCTGATCAAGAA-- 1660
QY 598 sGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp----- 613

```

Db      1661 -----ATTAAATCCTGGAGAGAAAAGAAATGGAATGCTTTGAAAATAT 1710
Qy      614 -TyrThrProGluIlePheSerPheGluSerThrThrGlyPheThrIleuTyrclyme 633
Db      1711 CCAGCTGCTTAAGAAATTAAGAACTTGAAGTAGAATGAATTAATTAATTAAGTACAA 1770
Qy      633 tleuTyrclymeProHieAspLeuGlnProGlyIleuTyrclymeProThrValIleuPheIleTy 653
Db      1771 GATGATTCCTCTCCCAATTTGACAGATCAAGAAAGTATCCCTTCAATTAATTAAGTGA 1830
Qy      653 rGlyGlyProGlnValGlnIleuValAsnAsnArgPheIleuValIleuTyrclymePheIle 673
Db      1831 TGGTGGTCCCTGACAGTCAAGTGTAAAGTCT-----GTATTTCTCTGT 1872
Qy      673 uAsnThrIleuAlaSerLeu-----GlyTyrclymeValIleuValIleuAsnArg 689
Db      1873 TAAATGATATCTTATCTTGAAGTAGAAGGAGTGGTCAATCCCTGGTGGATGATG 1932
Qy      689 gGlySerCysHieAspGlyIleuIleuPheGluGlyAlaPheIleuTyrclymeGlyGlnIle 709
Db      1933 AGGAACAGCTTCCAAAGTGACAACTCTCTATGCAAGTATCAAGAACTGGGTGTTTA 1992
Qy      709 eGluIleAspArgIleuValGlnIleuGlnIleuValIleuValIleuValIleuValIleu 729
Db      1993 TGAAGTTGAAGACCAAGATTACAGCTGTCAAGAAATTCATA---GAATGGGTTTCATTGA 2049
Qy      729 pleuAspArgValGlyIleuHieGlyThrSerTyrclymeTyrclymeSerLeuMetIleAla 749
Db      2050 TGAATAAAATAAGCCATATGGGCTGTCTATGAGATATGCTTCACTCACTGGCCCT 2109
Qy      749 uMetGlnArgSerAspIlePheArgValAlaIleuValIleuValIleuValIleuValIleu 769
Db      2110 TGCATCTGACATGCTGCTTTCAAAATGCTATAGCAAGGCTCCAGTCTCCAGCTGGGA 2169
Qy      769 ePheTyrclymeArgIleuTyrclymeGlyArgTyrclymeGlyArgPro-----AspGlnAsn 787
Db      2170 ATATTACGGCTGTCTACACAGAGATTCATGAGCTCTCCCAAGAAAGATGATATCT 2229
Qy      787 uGlnGlyTyrclymeGlySerValAlaMetGlnIleuValIleuValIleuValIleuValIleu 807
Db      2230 TGAGCATTAAGAAATTCACATGCTGTGCAAGAGCAATATTTTCAAGAAATGATAC-- 2287
Qy      807 nArgLeuIleuIleuHieGlyPheIleuAspGlnValIleuHieAspAlaIleuHieSerIle 827
Db      2288 ----TATCTTTCATCCACAGAAACAGCAATATATGCACTTTCAAAACTCAGACAA 2343
Qy      827 eIleuIleuSerPheIleuValArgAlaGlyIleuProTyrclymeGlnIleTyrclymeGln 847
Db      2344 GATTCCTAAAGCTGTGTTAATGACCAAGTGAATTCAGGCAATGTGTACTCTGACCA 2403
Qy      847 uArgHieSerIleuArgValProGluSerGly-----GluHieTyrclymeIleuHie 864
Db      2404 GAACCAACGCTTA-----TCCGCGCTGTCCAGAAACACTTATACACCAACAT 2451
Qy      864 uIleuHieTyrclymeGlnIle 870
Db      2452 GACCACTTCCCTAAAGCAG 2470

```

RESULT 6
US-08-619-280A-1
Sequence 1, Application US/08619280A
Patent No. 5767242

GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
Applicant: Retzig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York

```

COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION: DATA
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 638-3864
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 Base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-619-280A-1

Alignment Scores:
Pred. No.: 1,15e-41 Length: 2815
Score: 454.00 Matches: 177
Percent Similarity: 41.928 Conservative: 111
Best Local Similarity: 25.764 Mismatches: 278
Query Match: 9.664 Indels: 122
Gaps: 27
US-10-070-464-1 (1-882) x US-08-619-280A-1 (1-2815)
Qy      212 IleAlaPheIleHieSerAsnAspIleTrpIleSerAsnIleValThrArgGluIleArg 221
Db      692 TTAGCATATGCTATCAAAACAAATATCTATTGAAACAAAGACACAGAGATCCACCTTT 751
Qy      232 ArgLeuThrTyrclymeValIleuAsnGlnIleuAlaAsnMetGlnGluAspAlaArgSerAlaGly 251
Db      752 CAATTAACATTT-----AATGGAAGAGAAATAATTAATTAATGGA 793
Qy      252 ValAlaThrPheValIleuGlnIleuPhe-----AspArgTyrclymeGlyTyrclyme 269
Db      794 ATCCCAAGCTGGGTTTATGAGAGAGAAATGCTTCTCAAAATATGCT---CTGTGCTG 850
Qy      270 CysProIleuValGluThrThrProSerGlyIleuValIleuValIleuValIleuValIleu 289
Db      851 TCTCTCT-----AATGGAATAATTTTG---GCATATGGCGAA 883
Qy      290 AsnAspGluSerGluValGluIleuHieValIleuHieValIleuHieValIleuHieValIleu 308
Db      884 TTTAATGATTAAGATATACCACTTATGCTTATGCTTATGATGCGCATGAACAATATCT 943
Qy      309 ArgAlaAspSerPheArgTyrclymeProIleuTyrclymeThrAlaAspProIleuValIleuPhe 328
Db      944 AGAACAAATTAATATTCATCAACCAAGCTGAGCTAAGAAATCCGCTGTGCGGATA--- 1000
Qy      329 MetSerGluIleuMetIleuAspAlaGluIleuArgIleuIleuAspValIleuAspIleu 348
Db      1001 -----TTATATGATGATACCACTTACCTGCTAT 1030
Qy      349 IleGlnProPheGluIleuPheGluGlyValGluTyrclymeAlaArgAlaGlyTyrclyme 368
Db      1031 GTAGTCCCAAGAGTCTGTTCACCAATG-----ATAGCCTCAAGTATTAAT--- 1081
Qy      369 ProGluGlyTyrclymeAlaTrpSerIleuIleuAspArgSerGlnThrArgGluIle 388

```


Db 2344 GATTGCTAAGCTCTGTGTTAATGACACAGATGATTTCCAGGCAATGTGGTACTGACCA 2403
Qy 847 uArHisSer1LeaRyValProGluSerGly-----GluHisTyrGluLeuHisLe 864
Db 2404 GAACCCGCGCTTA-----TCCGCGCTGTCCAGCAACCACTTATACACCACAT 2451
Qy 864 uLeuHisTyrLeuGluGlu 870
Db 2452 GACCCACTTCTTAAGCAG 2470

RESULT 8
US-09-392-184-31/c
; Sequence 31, Application US/09392184
; Patent No. 6395889
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392,184
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; NAME/KEY: misc_feature
; LOCATION: (1)..(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31

Alignment Scores:
Pred. No.: 4,91e-37 Length: 612
Score: 404.00 Matches: 97
Percent Similarity: 69.28% Conservative: 9
Best Local Similarity: 63.40% Mismatches: 45
Query Match: 8.60% Indels: 4
Gaps: 0
DB: 4

US-10-070-464-1 (1-882) x US-09-392-184-31 (1-612)

Qy 710 Glu1LeaSPaSPGlnValGluGlyLeu-GlnTyrLeuAlaSerArgTyrAspPhe1Lea 729
Db 519 GAATTGTCCGATCAGCGTGGGNGGACTCCCAATATCTAATTCGAAATGATTCATGGA 460

Qy 729 Pleu-AspArgValGly1LeuHisGlyTyrPserTyrGlyTyrLeuSerLeuMetAlaL 749
Db 459 CTTAAATGCTGTGGCAATCCNGGCTG-TCCATAGAGATACCTCTCCCTGATGGCAN 401

Qy 749 eumErg1aRgSerAep1LePheArgValAla1LeaGlyAlaProValThrLeuTriP 769
Db 400 TAAATCAGAGCTCAGATATCNCAGGGTCT-ATTCTGGGGGCCAGATCAGCTTGA 342

Qy 769 1ePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGln 789
Db 341 TCTTCTATGAAACAGGAAACNCGGAGCTTANATGGGTCACTTNNCCAGNATGACAG 282

Qy 789 1TyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgL 809
Db 281 GCAATTACTTAGATCTGGCCATTCACNACAGAAAGTCCCTCTGNAACCAATNNGT 222

Qy 809 eulLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSer1LeuL 829
Db 221 TCTTCTCTTANATGGTTTCTGATTTAGGATTTCCATTTTCCANACCAAGNANNTAC 162

Qy 829 eulSerPheLeuValAlaGlyLysProTyrAspLeuGln1LeTyrProGlnGluArgH 849
Db 161 TTAGGTTTATGAGGCTGGGAAACCAATGANTTNCAGAACCAACCTCNGAGAAAGN 102

Qy 849 isSer1LeaRyValProGluSerGlyGluHisTyr 860
Db 101 CCAACANNAGAGTCTGATCGGAGCAATAT 67

RESULT 9
US-08-699-103B-1/c
; Sequence 1, Application US/08699103B
; Patent No. 6107462
; GENERAL INFORMATION:
; APPLICANT: Rine, Jasper D.
; APPLICANT: Hampton, Randolph
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: 16-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/002,581
; FILING DATE: 17-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 09272/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/322-5070
; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-699-103B-1

Alignment Scores:
Pred. No.: 5,29e-28 Length: 4982
Score: 339.50 Matches: 81
Percent Similarity: 52.07% Conservative: 32
Best Local Similarity: 37.33% Mismatches: 81
Query Match: 7.22% Indels: 23
Gaps: 7
DB: 3

US-10-070-464-1 (1-882) x US-08-699-103B-1 (1-4982)

Qy 646 TyrProThrValLeuPhe1LeTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPhe 665
Db 4941 TATCTCGTATTTTCTTTCATATGCGGACCAATTCACCAAGT----- 4894

Qy 666 LysGlyVal1LeTyrPheArgLeuAsn-----ThrLeuAlaSerLeuGlyTyr 681
Db 4893 -----GTCAAAACGTTTCCGTAGATTAAATGAAGTGTAGCTTCAACATTAACGCA 4840

Qy 682 ValValAlaVal1LeaSPaSPaRg1LeSerCyHisArgGlyLeuYserHeGluGlyAla 701
Db 4839 ATGTAGTGTGTGTGACGCGTGTGTACTGCTTCAAGAGTCAAGACTTTAGATCCCTT 4780

Qy 702 PheLysTyrLysMetGlyGln1LeGlu1LeaSPaSPGlnValGluGlyLeuGlnTyrLeu 721
Db 721


```
/
/      ZIP: 94025
/      COMPUTER READABLE FORM:
/      MEDIUM TYPE: Diskette
/      OPERATING SYSTEM: IBM Compatible
/      SOFTWARE: FASTSEQ for Windows Version 2.0
/      CURRENT APPLICATION DATA:
/      APPLICATION NUMBER: US/09/628,133
/      FILING DATE:
/      PRIOR APPLICATION DATA:
/      APPLICATION NUMBER: 08/699,103
/      FILING DATE:
/      ATTORNEY/AGENT INFORMATION:
/      NAME: Green, Grant D.
/      REGISTRATION NUMBER: 31,259
/      REFERENCE/DOCKET NUMBER: 09272/005001
/      TELECOMMUNICATION INFORMATION:
/      TELEPHONE: 650/322-5070
/      TELEFAX: 650/854-0875
/      INFORMATION FOR SEQ ID NO: 1:
/      SEQUENCE CHARACTERISTICS:
/      LENGTH: 4982 base pairs
/      TYPE: nucleic acid
/      STRANDEDNESS: single
/      TOPOLOGY: linear
/      MOLECULE TYPE: Genomic DNA
/
US-09-628-133-1

Alignment Scores:
Pred. No.: 5,29e-28      Length: 4982
Score: 339.50           Matches: 81
Percent Similarity: 52.07%      Conservative: 32
Best Local Similarity: 37.33%      Mismatches: 81
Query Match: 7.22%           Indels: 23
DB: 4                      Gaps: 7

US-10-070-464-1 (1-882) x US-09-628-133-1 (1-4982)
QY      646 TyrProThValIleuPheIleTyrgIyGlyProGlnValIleuValAsnAsnArgPhe 665
      |||||
DB      4941 TATCGTATTTCTTTCATATGCGGACCGAATTCACAAAGTT-----4894
QY      666 LysGIValLysTyRPhaArgLeuAsn-----ThleuAlaSerLeuGIYTyR 681
      |||||
DB      4893 -----GTCAAAACGCTTTCGATTAATGAAGTGAGTCTTCAACATTAACGCA 4840
QY      682 ValValValValIleAspAsnArgIySerCySHisArgIyLeuIyPheGIuGIyAla 701
      |||||
DB      4839 ATTGATGTTGTTGTCGATGCGTGTGCTGCTTCAAGGTCAGACCTTGAATCCCTT 4780
QY      702 PheIySTyLysMetGIyGlnIleGIuIleAspAsnGlnValIleGIyLeuGlnTyRleu 721
      |||||
DB      4779 GTTCGCGATAGGCTCGGTATTAACGAGCGCCGACCAAAATATCGCGCTTCTTATAT 4720
QY      722 AlaSerArgTyRAspPheIleAspLeuAspArgValGIyIleHISGIYTyRSeTyrgIy 741
      |||||
DB      4719 GGTTCCT--TTAACTTTGTTGATCCGCAAAAGATTTCCTTATTTGGTGGTCATACGGG 4663
QY      742 GlyTyRLeuSerLeuMetAlaLeuMetGlnArgSerAsp--IlePheArgValAlaIle 760
      |||||
DB      4662 GGGTACCTGACCTAATAAACTTTGAGAAAGATGGCGGAAGCATTTCAATAACGGGATG 4603
QY      761 AlaGIyAlaProValThIleuTrpIlePheTyRAspThrGIYTyRThGlnArgTyRmet 780
      |||||
DB      4602 TCAGTTGGCCGAGTAACGACGAGATTTTACGATTTCTGTTATACGAGAGATACATG 4543
QY      781 GlyHisProAspGlnAsnGlnIleGIYTyRLeuGIySerVal-----Ala 796
      |||||
DB      4542 CATACTCTCAAGAAACTTTGATGATGATCAAGGATCAATCAATGATCAATGCTACTGCT 4483
QY      797 MetGlnAlaGlnIyPheProSerGlnProAsnArgLeuLeuLeuLeuHISGIYPheLeu 816
      |||||
DB      4482 TTG-----GCACACAGCAATATGATTTTGTGATGACGAGAAACAGGA 4441
```

```
QY      817 AspGIuAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGIY 836
      |||||
DB      4440 GATGATACGTTACCTTCAAAATTCCTTAAGCTTTGGACCTTTGGATCTTAATGCT 4381
QY      837 ---LysProTyRAspLeuGlnIleTyRProGlnIuArgHisSerIleArg 852
      |||||
DB      4380 GTGAAATATGACGCTCAGCTCTTCTGACGATCATATGATTAAGA 4330

RESULT 12
US-09-221-017B-646/C
/      Sequence 646, Application US/09221017B
/      Patent No. 6444799
/      GENERAL INFORMATION:
/      APPLICANT: Ross, Bruce C.
/      TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
/      NUMBER OF SEQUENCES: 1120
/      CORRESPONDENCE ADDRESS:
/      ADDRESSEE: MORRISON & FOERSTER
/      STREET: 755 PAGE MILL ROAD
/      CITY: Palo Alto
/      STATE: CA
/      COUNTRY: USA
/      ZIP: 94304-1018
/      COMPUTER READABLE FORM:
/      MEDIUM TYPE: Diskette
/      OPERATING SYSTEM: Windows
/      SOFTWARE: FASTSEQ for Windows Version 2.0b
/      CURRENT APPLICATION DATA:
/      APPLICATION NUMBER: US/09/221,017B
/      FILING DATE: 23-DEC-1998
/      CLASSIFICATION:
/      PRIOR APPLICATION DATA:
/      APPLICATION NUMBER: PP1182
/      FILING DATE: 31-DEC-1997
/      PRIOR APPLICATION DATA:
/      APPLICATION NUMBER: PP1546
/      FILING DATE: 30-JAN-1998
/      PRIOR APPLICATION DATA:
/      APPLICATION NUMBER: PP2911
/      FILING DATE: 09-APR-1998
/      PRIOR APPLICATION DATA:
/      APPLICATION NUMBER: PCT/AU98/01023
/      FILING DATE: 10-DEC-1998
/      ATTORNEY/AGENT INFORMATION:
/      NAME: Monroy, Gladys H
/      REGISTRATION NUMBER: 32,430
/      REFERENCE/DOCKET NUMBER: 27340-20021.00
/      TELECOMMUNICATION INFORMATION:
/      TELEPHONE: 650-813-5600
/      TELEFAX: 650-494-0792
/      TELEX: 706141
/      INFORMATION FOR SEQ ID NO: 646:
/      SEQUENCE CHARACTERISTICS:
/      LENGTH: 657 base pairs
/      TYPE: nucleic acid
/      STRANDEDNESS: double
/      TOPOLOGY: circular
/      MOLECULE TYPE: DNA (genomic)
/      HYPOTHETICAL: NO
/      ANTI-SENSE: UNKNOWN
/      ORIGINAL SOURCE:
/      ORGANISM: PORPHYROMONAS GINGIVALIS
/      FEATURE:
/      NAME/KEY: misc_feature
/      LOCATION: 1...657
/      US-09-221-017B-646

Alignment Scores:
Pred. No.: 2.95e-28      Length: 657
Score: 328.50           Matches: 71
Percent Similarity: 55.26%      Conservative: 34
```

Best Local Similarity: 37.37% Mismatches: 78
 Query Match: 6.99% Indels: 7
 DB: 4 Gaps: 3

US-10-070-464-1 (1-882) x US-09-221-017B-646 (1-657)

```

QY 614 TTTTCTTCTGATGTTTCACTATCAAAACCAATCGCTTGAACATGATGCTCG 507
DB 566 TTAACCCGAGAGGTTTCACTATCAAAACCAATCGCTTGAACATGATGCTCG 507
QY 634 LeuTyrLysProHisAspLeuGlnProGlyLysTyrProThrValLeuPheIleTyr 653
DB 506 ATCGGAACCTATTGATTTCCATCCCTCCGCACTATCTGCTGATGTAACGATAT 447
QY 654 GtlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 673
DB 446 ACCGGTCCCACTCCAGCAGGATTGATCGCTAT-----TCATTGATTGGGA 396
QY 674 AsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgLysSerCysHis 693
DB 395 CACTACCTTGCTGATCGAAGGTTACGTCGCGCATGTGGATGGCGCTGGCAGCGGTGCT 336
QY 694 ArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAsp 713
DB 335 CCGCGCGAAGATGCGCAAGTGTACTCATGCACTGCGTGTATTCCAAAGCGATGAT 276
QY 714 GlnValGlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgVal 733
DB 275 CAGATAGCAGCGCCACTGCTCTATA--GGACAGCTGCCCTATGTGATGACACTGCTATTC 219
QY 734 GtlyIleHisGlyTyrPsetTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 753
DB 218 GGCATATAGGGGTTGAGGCTATAGCGCGCTATACCACTATAGATTTGTTGCTGGGGAAT 159
QY 754 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThr 773
DB 158 GGTACATTCAAAGCGGAGTAGCCGCTGCTCTGTCGCGACAGCTGGCGTTCATCAGATTG 99
QY 774 GtlyTyrThrGlnArgTyrMetGlyHisProAspGlnAsnGlnGlyTyrTyrLeuGly 793
DB 98 GTTTACACCGAAGCTTACGTGCTACACCCAAAGAAATGCTTCCGATTAACAAGATGCT 39
QY 794 Ser-----ValAlaMetGlnAlaGln 800
DB 38 TCTGCTTGTGATGTGGCAAGCAATTACAA 9

```

RESULT 13
 US-09-221-017B-253
 ; Sequence 253, Application US/09221017B
 ; Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSS, Bruce C.
 TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 PAGE MILL ROAD
 CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/221,017B
 FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1182
 FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546
 FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP2911
 FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU98/01023
 FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:
 NAME: Monroy, Gladys H.
 REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 253:
 SEQUENCE CHARACTERISTICS:

LENGTH: 543 base pairs
 TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN
 ORIGINAL SOURCE:

ORGANISM: PORPHYROMONAS GINGIVALIS
 FEATURE:

NAME/KEY: misc feature
 LOCATION: 1...543

US-09-221-017B-253

Alignment Scores:

Score: 2,42e-28 Length: 543
 Percent Similarity: 328.00 Matches: 63
 Best Local Similarity: 58.82% Conservative: 37

Query Match: 6.98% Mismatches: 66
 DB: Indels: 4
 Gaps: 2

US-10-070-464-1 (1-882) x US-09-221-017B-253 (1-543)

```

QY 676 LeuAlaSerLeuGlyTyrValValValIleAspAsnArgLysSerCysHisArgGly 695
DB 43 ATGGCAGCAAGAGGCTATGCGCTTACGGGATAGCGCGGATCTCCATTAAGAGG 102
QY 696 LeuLysPheGlnGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAsnVal 715
DB 103 GCTGCTTTCAGCAGGTTATTCATGCTGTTGGGCGACAGCGATGCCATCAAGATG 162
QY 716 GtlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 735
DB 163 TCGGGTGTGATTTCTCTCAAGCCAA--TCATGGTGGAATGCCGATAGATGAGAGTA 219
QY 736 HisGlyTyrPsetTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 755
DB 220 CATGGCTGAGCTATGTTGGCTTCTTATGACTACGAATCTGATCTTAACGACCGCGATGTC 279
QY 756 PheArgValAlaIleAlaGlyValaProValThrLeuTyrIlePheTyrAspThrGlyTyr 775
DB 280 TTCAAAAGTCGAGTAGCCGCGCGGCTGTCTCATAGACTGAATCGATATAGATTAATGATG 339
QY 776 ThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlyTyrTyrLeuGlySerVal 795
DB 340 GGTGAGCGTTATTTGATGCGCCACAGAAATCCGAAAGATACGATGCTGCCAACCTG 399
QY 796 AlaMetGlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPhe 815
DB 400 CTCAAACGAGCCGATGATCG-----AAAGCAGCATTAATCTGATTAATGAGCG 450
QY 816 LeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAla 835

```

Db 451 ATGCATCGGTCGTATGCGACAGCATTCCTTTCTTGATGCTGTGGTGAAGCA 510
Qy 836 GlySerProTyrAspLeuGlnIleTyrPro 845
Db 511 CGCACTTACCTGACTCTTACGTCTATCCG 540
RESULT 14
US-09-016-080-2
Sequence 2. Application US/09016080
Patent No. 6133012
GENERAL INFORMATION:
APPLICANT: Ishikawa, Kazuhiko
APPLICANT: Matsui, Ikuo
APPLICANT: Ishida, Hiroyasu
APPLICANT: Kozugi, Yoshiyuki
APPLICANT: Higuchi, Kazuhiko
TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
FILE OF INVENTION: THE SAME
FILE REFERENCE: 07898/022001
CURRENT APPLICATION NUMBER: US/09/016,080
EARLIER FILING DATE: 1998-01-30
EARLIER APPLICATION NUMBER: JAPAN 18381/1997
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 1896
TYPE: DNA
ORGANISM: Pyrococcus horikoshii
FEATURE:
NAME/KEY: CDS
LOCATION: (1) .. (1896)
US-09-016-080-2
Alignment Scores:
Pred. No.: 4,38e-22 Length: 1896
Score: 282.00 Matches: 151
Percent Similarity: 38.29% Conservative: 91
Best Local Similarity: 23.89% Mismatches: 242
Query Match: 6.00% Indels: 148
Gaps: 33
US-10-070-464-1 (1-882) x US-09-016-080-2 (1-1896)
Qy 276 ThrProSerGlyGlyIleLeuArgIleLeuTyrGlnGluAsnAspGluSerGluVal 295
Db 199 TCTCCAAATGGGAAGCTTATAGCATTTACTCAAGGGAATGAAGAAAGAAATCA 238
Qy 296 GlnIleIleHisValThrSerProMetLeuGluThrArgAlaAspSerPheArgTyr 315
Db 259 GACCTTACCTGATTCACACGAGATGGGAGAGGCCAGACTTTTACCAAGTTCAATATAC 318
Qy 316 ProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMet----- 333
Db 319 -----GGGATTAAGAACTGCGCTTATCCGAGATGGGAAAGCTATAGCCGTGTT 369
Qy 334 -----IleAspAlaGluGlyArgIleIleAspValIleAspLysGluLeuIleGlnPro 351
Db 370 ACCCCTATACCTGTGGAAGAAAGGAATGAT-----GACGTTCACATTAATAGGAA 423
Qy 352 PheGlnIleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTyrPthrProGluGly 371
Db 424 ATACCAATTCGTGTTTATGAGATT-----GGCTGATC---TAGCGA 462
Qy 372 LysTyrAlaIleTyrSerIleLeuLeuAsp-----ArgSerGlnThrArgLeu----- 386
Db 463 AAAAGAAAGCTGTCTACCTTGTGACGTGGAGACGGGAAAGAAAGACCTAATCTCA 522
Qy 387 -----GlnIleValIleLeuSerProGluLeuPheIleProValGlu 400
Db 523 AAGCACTTAATGTGATCAGATTAAGGTTCCACACAGCGTAGACTTATCTTACAGGCCCA 562
Qy 401 AspAsp-----ValMetGluArgGln----- 407

Db 583 GAGATATGGAAGAAAGAACTGTGATATCCGATCTTTAGCTCTCCGGAATAGAAAGTT 642
Qy 408 -----ArgLeuIleGluSerValProAspSerValThrPro 419
Db 643 AGAAGCTGACCCGAGGAGATGAGATGACTTCTCCCTTCGTATGACGGAAC 702
Qy 420 LeuIleIleTyrGlnGluThrAsp-----IleTyrIleLeuIleHisAspIlePhe 437
Db 703 TTGCTACTTAAGCTTACCTTTAGAAAGGGAATCCACCAAGCCGAC---ATTAC 759
Qy 438 HisValPheProGluSerHisGluGluIleGluPheIlePheAlaSerGluCysLys 457
Db 760 CACTAGCATCC-----AAG 774
Qy 458 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 477
Db 775 ACAGAG-----GACCTTAAGAGCTCACAAAGATTATACAGAAACGCTTACATCC 828
Qy 478 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 497
Db 829 TTAAAC-----TCCGATGTTGAGGAAGTCAAGAGCCGAGCTGTG 870
Qy 498 IleThrSerGlyLysThrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 517
Db 871 TACAGAGAGGG---TGG----- 885
Qy 518 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisLeuTyr 537
Db 886 -----ATCACTATATGCGCAACG---GATGCGCCT---AGGCAAACTCTTT 927
Qy 538 ValValSerTyrValAsnProGlyValThrArgLeuThr-----AspArgGlyTyr 555
Db 928 AGGTCACACTTA-----GATGAAAGATTGAAAGGATTAAGCTGAGATAGAACGTT 981
Qy 558 SerHisSerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGln 575
Db 982 GAA-----AGCTTGATATAGGGAATTAATAGCTTTC 1014
Qy 576 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 595
Db 1015 ACGGCTCAAGATCCTTAACCCCACTGAGCTGTATATACAGGATGAAAGCAAGAAC 1074
Qy 596 LysThrLysGluPhe-----TrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp 613
Db 1075 AAGCTTACCGACTTAAACAATGG-----ATTAACGGT 1107
Qy 614 TyrThr-----ProProGluIlePheSerPheGluSerThrArgLysPheThrLeuTyr 631
Db 1108 TACACCTTTCAAAACCTGAACACTTTAAGTTAAAGCAAGTACGCGGTGAATATGAT 1167
Qy 632 GlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPhe 651
Db 1168 GCCTGGATATGAAACCGGTGAACCTTACAGAAAGGAAGATATACAGCTATTTTACAG 1227
Qy 652 IleTyrGlyLysProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPhe 671
Db 1228 ATCCAGCGTGTCTTAAACCGCTTAC-----GATTACGCTTTATG 1269
Qy 672 ---ArgLeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGly 690
Db 1270 CACGAGTTCACGTTTAAACCTTAAAGGCTTGCTGTATATTCCTCAATCTTACAGGG 1329
Qy 691 SerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlnIleGlu 710
Db 1330 AGCGATGGCTACGAGAGGAGTTC---GGCGATATTAAGGGGACATATGAGGAGAGGAT 1386
Qy 711 IleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeu 730
Db 1387 TACAGAGATTTAATGAGATGATGATGAAGCTTAAGAGATTTACTTCAATAGTGG 1446
Qy 731 AspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMet 750
0

Db 1447 GAAAGCTAGAGTTACCGGGGTTCTATGGTGCCTTCATGACGAAGTGG---ATAGTC 1503
Qy 751 GlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThiLeuTrpIlePhe 770
Db 1504 GCACATATACCAAGGTTCAAGCCCGTGAACCCAGATCAATTCATTAATGGAATTAAGC 1563
Qy 771 Tyr-----AspThrGlyTyrThrGluArgTyrMetGlyHisProAspGln--- 785
Db 1564 TTCTTCGGGACACGAGTATAGGTTAT-----TACTTGGCT---CCAATCAATA 1611
Qy 786 -----AsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAla 799
Db 1612 GGAAGATATCCCTGGAGCAACTTGGAGAGGTATGGAAGAAAGAGCCCATTAAGTACGCT 1671
Qy 800 GluIysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsn 819
Db 1672 -----CCCAAGCTTGAACCTCCCTGCTTATATACCTGTAACCAAGACTACAG 1722
Qy 820 ValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIysProTyr 839
Db 1723 TGTGGCTTCCCGAGGACTTGCACACTTTCATATCCCTAAATACCTGGGAGAGAGTT 1782

RESULT 15
US-09-389-681-428
; Sequence 428, Application US/09389681A
; Patent No. 6518237
; GENERAL INFORMATION:
; APPLICANT: Yuuni, Jiang
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C3
; CURRENT APPLICATION NUMBER: US/09/389,681A
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 428
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-389-681-428

Alignment Scores:
Pred. No.: 1.24e-22 Length: 535
Score: 278.50 Matches: 64
Percent Similarity: 55.23% Conservative: 31
Best Local Similarity: 37.21% Mismatches: 64
Query Match: 5.93% Indels: 13
DB: 4 Gaps: 4

US-10-070-464-1 (1-862) x US-09-389-681-428 (1-535)

Qy 637 ProHisAspLeuGlnProGlyIlyValYsTyrProThrValLeuPheIleTyrGlyIyPro 656
Db 15 CCTCCTCAATTGACAGATCAAGAAAGTATCCCTGCTAATTCACAGTATGATGTGTCCC 74
Qy 657 GlnValGlnLeuValAsnAsnArgPheIysGlyValIysTyrPheArgLeuAsnThrLeu 676
Db 75 TGCAGTCAAGAGTGAAGTCT-----GTAATTCCTGTTAATGAGATA 116
Qy 677 AlaSerLeu-----GlyTyrValValValValIleAspAsnArgGlySerCys 692
Db 117 TCTTATCTTGCAAGTGAAGAGGATGTCATGCTTGATGATGTCGAGAACAGCT 176
Qy 693 HisArgGlyLeuIysPheGluGlyAlaPheIysTyrIlyMetGlyGlnIleGluIleAsp 712
Db 177 TTCAGAGTGAACAATCCTCTATGCAAGTATGCAAGAGCTGGGTGTTATGAAAGTTGAA 236

Qy 713 AspGlnValGluGlyLeuGlnIlePheValAspArgTyrAspPheIleAspLeuAspArg 732
Db 237 GACCGATTTACAGCTGTGCAAGAAATTCATA---GAAATGGGTTTCATTCATGATGAAAAAGCA 293
Qy 733 ValGlyIleHisGlyTrpSerTyrGlyIyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
Db 294 ATAGCATATAGGGGCTGGTCCCTATGAGAGATACGTTTCATCACTGGCCCTTGATCTGGA 353
Qy 753 SerAspIlePheArgValAlaIleAlaGlyAlaProValThiLeuTrpIlePheTyrAsp 772
Db 354 ACTGGCTTTTCAAAATGCTGTATACAGTGCCTCCAGTCTCCAGCTGGGAATATTAACGGC 413
Qy 773 ThrGlyTyrThrGluArgTyrMetGlyHisPro-----AspGlnAsnGluGlnIyTyr 790
Db 414 TCTGCTACACAGAGATTCATGGGCTCTCCCAACAAGATGATATCTTGAGACATAT 473
Qy 791 TyrLeuGlySerValAlaMetGlnAlaGlyIysPhe 802
Db 474 AAGATTCACCTGTGATGCGCAAGACAGAAATTTTC 509

Search completed: October 16, 2003, 03:30:10
Job time : 190.181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 16, 2003, 00:05:32 (Search time 659.423 Seconds
(without alignments))

3512.533 Million cell updates/sec

Title: US-10-070-464-1

Perfect score: 4700

Sequence: 1 MAAMETEDLGVEIFETADC.....HLHYLQENLGRIRALKYI 882

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh

-Q=/cg2_1/USFTO.spool/US10070464/runat_15102003_113555_24924/app.query.fasta_1.2652

-DB=published Applications NA -QFMT=fastap -SUFFIX=rmdb -MINMATCH=0.1

-LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bl0sum62

-TRANS=human0.ccl -LIST=45 -DOCALLIGN=200 -THR SCORE=ptc -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US10070464@CGN.1.1.845@runat_15102003_113555_24924

-NCPU=6 -ICPU=3 -NO MAP -LARGESQERY -NEG SCORES=0 -WAIT -DSRBLCK=100

-FONLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-LONPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cg2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cg2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cg2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cg2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cg2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cg2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cg2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cg2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cg2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cg2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*

11: /cg2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cg2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*

13: /cg2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*

14: /cg2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cg2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*

16: /cg2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*

17: /cg2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4700	100.0	2649	12	US-10-054-776-1

2	4700	100.0	2649	12	US-10-170-789-39	Sequence 39, Appl1
3	4700	100.0	2671	10	US-09-976-674-2	Sequence 2, Appl1
4	4700	100.0	3143	12	US-10-170-789-37	Sequence 37, Appl1
5	4680	99.6	4829	10	US-09-976-674-12	Sequence 12, Appl1
6	4385.5	93.3	4685	10	US-09-976-674-22	Sequence 22, Appl1
7	4385	87.1	4676	10	US-09-976-674-20	Sequence 20, Appl1
8	4092.5	77.9	4309	10	US-09-976-674-8	Sequence 8, Appl1
9	3661.5	77.9	4309	10	US-09-976-674-14	Sequence 14, Appl1
10	2870	61.1	2617	10	US-09-976-674-1	Sequence 4, Appl1
11	2870	61.1	4219	10	US-09-976-674-28	Sequence 28, Appl1
12	2870	61.1	4302	10	US-09-976-674-24	Sequence 24, Appl1
13	2820.5	60.0	4180	10	US-09-976-674-36	Sequence 36, Appl1
14	2820.5	60.0	4263	10	US-09-976-674-34	Sequence 34, Appl1
15	2649	56.4	4076	10	US-09-976-674-32	Sequence 32, Appl1
16	2649	56.4	4159	10	US-09-976-674-30	Sequence 30, Appl1
17	2638	56.1	2801	13	US-10-098-841-100	Sequence 100, Appl
18	2599.5	55.3	4037	10	US-09-976-674-40	Sequence 40, Appl1
19	2599.5	55.3	4120	10	US-09-976-674-38	Sequence 38, Appl1
20	2476.5	52.7	3262	13	US-10-098-841-83	Sequence 83, Appl1
21	1400	29.8	2079	13	US-10-044-090-843	Sequence 843, Appl
22	1391	29.6	2411	10	US-09-976-674-26	Sequence 26, Appl1
23	1278	27.2	1356	10	US-09-976-674-10	Sequence 10, Appl1
24	1026.5	21.8	832	10	US-09-976-674-18	Sequence 18, Appl1
25	805	17.1	502	11	US-09-918-995-19585	Sequence 19585, A
26	665.5	14.2	561	11	US-09-764-891-877	Sequence 877, Appl
27	529	11.3	3407	12	US-10-423-714-5	Sequence 5, Appl1
28	529	11.3	3407	13	US-10-002-593-5	Sequence 5, Appl1
29	529	11.3	3407	14	US-10-165-603-6	Sequence 6, Appl1
30	524	11.1	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
31	524	11.1	4835	14	US-10-165-603-5	Sequence 5, Appl1
32	505	10.7	2130	14	US-10-156-761-3131	Sequence 3131, Ap
33	505	10.7	9025608	14	US-10-156-761-1	Sequence 1, Appl1
34	464	9.9	620	10	US-09-976-674-16	Sequence 16, Appl1
35	461	9.8	2388	12	US-09-870-133-3	Sequence 3, Appl1
36	461	9.8	2388	14	US-10-160-501-6	Sequence 6, Appl1
37	461	9.8	2583	10	US-09-976-674-6	Sequence 6, Appl1
38	461	9.8	3238	12	US-09-870-133-1	Sequence 1, Appl1
39	461	9.8	3238	14	US-10-160-501-4	Sequence 4, Appl1
40	461	9.8	4541	10	US-09-976-674-42	Sequence 42, Appl1
41	456	9.7	2366	12	US-10-101-510-683	Sequence 683, Appl
42	456	9.7	2788	12	US-10-269-909-22	Sequence 22, Appl1
43	456	9.7	3128	14	US-10-198-845-1171	Sequence 13171, A
44	455	9.7	3224	12	US-10-240-965-117	Sequence 117, App
45	454	9.7	2814	10	US-09-962-832-108	Sequence 108, App

ALIGNMENTS

RESULT 1
US-10-054-776-1
; Sequence 1, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: OG1042US
; CURRENT APPLICATION NUMBER: US/10/054, 776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2649)
US-10-054-776-1
Alignment Scores: 0
Pred. No.: 4700.00
Length: 2649
Matches: 882

Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 12
Gaps: 0

US-10-070-464-1 (1-882) x US-10-054-776-1 (1-2649)

QY 1 MetAlaAlaIleMetGluThrGluGluLeuGlyValGluIlePheGluThrAlaAspCys 20
DB 1 ATGCACACAGCAATGGAAGAAACAGAACAGCTGGGTGTGGATATTTGAAACCGCGACGTG 60
QY 21 GluGluAsnIleGluSerGluAspAspProlyserGluGluProPheTyrValGluArgTyr 40
DB 61 GAGAGAAATATTGATATCACAGGATCGGCTTAATTGGAGCCCTTTTATGTTGACGGGTAT 120
QY 41 SerTyrSerGluLeuGlyValLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
DB 121 TCCTGGAGTCAAGCTTTAAAAAGCTGCTGGATCCAGAAATATCATCGCTACATGATG 180
QY 61 AlAlaValAlaProHisAspPheMetCysValLysArgAsnAspProAspGlyProHisSer 80
DB 181 GCTAAGGACCAATGATTCATGTTGTGTAAGAGATGATCCAGATGGAACCTCATTTCA 240
QY 81 AsparGlyIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 241 GACGAAATCTATTACCTTGGCATGCTGTGTAGAAACAGAAATATCATGTTTATTTCT 300
QY 101 GluIleProIleThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu 120
DB 301 GAAATTTCCAAACATCATCATATAGAGACAGCATTTAATGCTCTCTGGAGACCTTTTGG 360
QY 121 AspleuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB 361 GATTTTTTTCAGGACACCTGAGCTATGGAATGATTTCTCGAAGAAAGAACTATTAGA 420
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 421 GAAAGAAACCCATTTGGAAACAGTGGAAATTCCTTTCATCATTTATCACAAAGAAAGTGA 480
QY 161 ThrPheLeuPheGluAlaGlySerGlyLysTyrHisValLysAspGlyGlyProGlnGly 180
DB 481 ACATTTCTGTTTCAAGCGGTAGTGAATTTATCATCGTAAAGATGGAGGCCCAAGAA 540
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 541 TTTAAGCAACACCTTTAAGGCCCAATCTAGTGAAGAACTAGTGTCTCCCAATATCGAGTG 600
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 601 GATCAAAAATTATGCCCTCGCTGATCCAGACTCGATTTGCTTTTATACATAGCAACGATTT 660
QY 221 TrrIleSerAsnIleValThrArgGluGluArgLysLeuThrTyrValHisAsnGluLeu 240
DB 661 TGGTATCTAATCATCGTAAACAGAGAAAGAGACTCACTTATGTGCACAAATGAGCTA 720
QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValIleLeuGlnGlu 260
DB 721 GCCAACAATGGAAGAAATGCGAGATCAGCTGAGATCGCTACTTTGTTCTCAAGAAAGAA 780
QY 261 PheAspArgTyrSerGlyTyrTrrTrrCysProLysAlaGluThrThrProSerGlyGly 280
DB 781 TTTGATATGATATTCTGGCTATTGGTGTGCTCCAAAGCTGAACAACTCCCAAGTGGTGT 840
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGlyValGluIleIleHisVal 300
DB 841 AAAATTTCTTAATTTCTAATGAAAGAAATGATGAATCTGAGTGGAAATATTATCATGTT 900
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB 901 ACATCCCTTATGTTGGAAACAGAGAGGACATTCATTCCTTATCTCTAAACAGGTACA 960
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340

DB 961 GCAATCTTAAGCTACCTTTAAGATGTCAGAAATATGATGATGCTGAAGAAAGATC 1020
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB 1021 ATAGATGTCATAGTAAAGAACTAATTCACACCTTTTGAGATTTGATTTGAAAGAGTTGAA 1080
QY 361 TyrIleAlaArgAlaGlyTrrThrProGluGlyLysTyrValAspSerIleLeuLeuAsp 380
DB 1081 TATATTGCCAGAGCTGATGAGACTCTCGAGGAAATATGCTGGTCTCATCTCATGAT 1140
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1141 CGCTCCAGACTCGCTCGAGATGATGTTGATCTCACCTGAATTTATTTATCCAGTGA 1200
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1201 GATGATGTTATGGAAGAGCAGAGACTCATTTAGTCACTGCTGATTTCTGTGAGCCCACTA 1260
QY 421 IleIleTyrGluGluThrThrAspIleTrrIleAsnIleHisAspIlePheHisValPhe 440
DB 1261 ATTATCTATGAGAAACAGACATCTGATTAATTTCCATGACATCTTTTCATGTTT 1320
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
DB 1321 CCCCAAGTCAAGAGAGAAATGAGTTTATTTTCTCTGAAATCCAAACAGTTTC 1380
QY 461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
DB 1381 CGTCATTTATACAAATATTCATCATCTTTTAAAGAAAGCAAAATTAACATCCAGTGGT 1440
QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB 1441 GGGCTGCTCTCCAAATGATTTTCMACTGTCCTATCAAAAGAGATAGCAATTAACAGT 1500
QY 501 GlyIleTrrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArg 520
DB 1501 GGTGAATGGGAAGTCTTGGCCGAGATGATCTTAATATCCAAAGTTGATGAAGTGAAGG 1560
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeuTyrValIleSer 540
DB 1561 CTGGTATATTGTAAGGACCAAAAGACTCCCTTTAGCATGACCTGTACGTAGTCACT 1620
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgLysTyrSerHisSerCys 560
DB 1621 TACGTAAATCTGAGAGGAGGACAAAGCTGACGATCACTCACTCTTCTTGGTGC 1680
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsnProHisCys 580
DB 1681 ATCAGTACGACACTGTGCTCTTTATTAAGTATGATATGTAACCAAGAAATCCACACTGT 1740
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
DB 1741 GTGTCCCTTACAAAGTATCAAGCTCGAAGATGACCACCACTGCAAAACAAAGAAATTT 1800
QY 601 TrrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
DB 1801 TGGGCCACCAATTTGGATTGACAGAGCTCTTCTGACATATCTCTCCAGAAATTTTC 1860
QY 621 SerPheGluSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
DB 1861 TCTTTTGAAGTACTACTGATTTTACATTGATGAGATGCTCTACAGCCCTCATGATCTTA 1920
QY 641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
DB 1921 CAGCTCGAAGAAAGAAATATCTACTAGTGTCTTATATATATGATGCTCTGAGTGA 1980
QY 661 ValAsnAspArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
DB 1981 GTGAATATATGTTTAAAGAGTCAGATATTTCCGCTGAATATCCCTAGGCTCTTAAGT 2040
QY 681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLysLeuLysPheGluGly 700
DB 2041 TATGTGTTGTAATGATGATGACCAAGGATCTGTACCGAAGGCTTAAATTTGAAGGC 2100

```

QY 701 AlaPheLysTyrIleuSerMetGluGlnIleGluLeuAspArgValGluGlyLeuGlnTyr 720
DB 2101 GCGTTTAAATATATAATGGGTCAATATGAAATTCACATACAGTGGAAAGATCTCCAAATAT 2160
QY 721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyr 740
DB 2161 CTAGGCTCTCGATATGATTTTCATTGACTAGATGCTGTGGGCATCCACGGCTGTCTAT 2220
QY 741 GlyGlyTyrIleuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
DB 2221 GGAGGAATACCTCTCCGATGGCAATATGACAGAGTCAGATATCTTCAGGGTTGCTATT 2280
QY 761 AlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
DB 2281 GCTGGGGCCCGACGACCTCTGTGATCTTCTATATACAGGATACAGGAACGTTATATG 2340
QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrIleuGlySerValAlaMetGlnAlaGlu 800
DB 2341 GGTCAACCTGACGCAATGAAACAGGGCTATTACTTAGATCTGTGGCCATGCAAGCAAGA 2400
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
DB 2401 AAGTTCCCTCTGAAACCAATCGTTTACTGCTTACATGGTTTCCGAGATGAAATGTC 2460
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
DB 2461 CATTTTGACATACCACTATATTTACTAGTTTTTTATGAGGGCTGAAAGCCATATGAT 2520
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
DB 2521 TTACAGATCTATCTCTGAGAGACACAGCATAGAGTCTGAAATCTGGAGAACATTTAT 2580
QY 861 GluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLys 880
DB 2581 GAACGTGATCTTTGGACTACCTTCAAGAAACCTTGATACGATTTGCTGCTTAA 2640
QY 881 ValIle 882
DB 2641 GTGATA 2646

```

```

; PRIOR APPLICATION NUMBER: US 60/226,740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-170-789-39

Alignment Scores:
Pred. No.: 0 Length: 2649
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 12

US-10-070-464-1 (1-882) x US-10-170-789-39 (1-2649)
QY 1 MetAlaAlaIleMetGluThrGluGlnIleGluValGluIlePheGluThrAlaAspCys 20
DB 1 ATGGACAGCAGCATGGAACAGAACAGCTGGGTGATATTTGAAATCTGGAGCTGT 60
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
DB 64 GAGGAGAAATATTGAATACACAGATCGGCCCTAAATTGGAGCTTTTATGTTGACGGTAT 120
QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
DB 121 TCTTGAGTCAGCTTAAAGAGCTGCTGGATCCAGAAATATATATGCTCATGATG 180
QY 61 AlaIleAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
DB 181 GCTAAGGACCAATATATTTTCATGTTTGTGAAAGAAATCATCATGATGACCTCATTTCA 240
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB 241 GACAGATCTATTAACCTTGGCATGTCTGGTGAAGAACAGAAATATACACTGTTTATTTCT 300

```

QY 101 GluileProlyserThrileAsnArgAlaIvalleuMetleuSerTrpIysProleuLeu 120
 |||||
 Db 301 GAAATTCCAAACATATCAATAGACAGAGCTTAAATGCTCTTGGAAAGCCCTTTTG 360
 QY 121 AspleuPheGlnAlaThrleuAspTyrglyMetTySerArgGluGluLeuLeuArg 140
 |||||
 Db 361 GATCTTTTTCAGGCAACACTGAGATATGGAATGTATTCGAGAAAGAAATCTTTAGA 420
 QY 141 GluArgLysArglleGlyThrValGlylleAlaSerTyraPtyrhisGlnGlySerGly 160
 |||||
 Db 421 GAAGAAGAAACGATGTGAAACAGTCGGAATGCTCTTACATATTCACCAAGGAAGTGA 480
 QY 161 ThrPheleuPheGlnAlaGlySerGlylleTyrhisValLysAspArglyProGlnGly 180
 |||||
 Db 481 ACATTTCTGTTTCAAGCCGAGTGAAGATTTATACCTTAAAGATGAGAGGCCCAAGGA 540
 QY 181 PheThrGlnGlnProleuArgProAsnleuValGluThrSerCysProAsnIleArgMet 200
 |||||
 Db 541 TTATCGCAACACCTTTAAGGCCCAATCTAGTGAATAGTTGTCCCAACATACGAGATG 600
 QY 201 AspProLysleuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
 |||||
 Db 601 GATCCAAATTAATGCTGCTGATCCAGACTGAGATTCCTTTTAACATAGCAACGATATT 660
 QY 221 TrpIleSerAsnIleValThrArgGluGluArgArgleuThrTyraIleHisGlnLeu 240
 |||||
 Db 661 TGGATATCTTAACATCTGAACCGAAGAAAGAGACTCACTTATGTGCAATGAGCTTA 720
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValleuGlnGlu 260
 |||||
 Db 721 GCCAACTGMAAGAAAGATGCGAGATCAGCTGAGCTGCTACCTTGTCTCCAGAGAA 780
 QY 261 PheAspArgTySerArglyTyrrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
 |||||
 Db 781 TTTGATGTGAATTCGTGCTATGTGGTGTCCAAACCTGMAACCTCCACAGGAGGT 840
 QY 281 LysIleleuArgIleleuTyrgluGluAsnAspGluSerGluValGluIleHisVal 300
 |||||
 Db 841 AAAATCTTGAAATTCATATGAAAGAAATGATGATCTGAGGTGGAATTTATTCATGTT 900
 QY 301 ThrSerProMetleuGluThrArgArgAlaAspSerPheArgTyrrProLysThrGlyThr 320
 |||||
 Db 901 ACATCCCTATGTGGAACCAAGAGGCGAGATTCATTCCTGCTTAAACACAGTACA 960
 QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
 |||||
 Db 961 GCAAATCTTAAAGTCACTTTAAGATGTCAGAAATTAATGATGTGTGAAGGAAGATC 1020
 QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleleuPheGluGlyValGlu 360
 |||||
 Db 1021 ATAGATGTCAATAGTAAAGAACTAAATTCACCTTTGAGATTCATTTGAAGAGATTGAA 1080
 QY 361 TyrrIleAlaArgAlaGlyTrpThrProGluGlyLysTyraIleTrpSerIleleuLeuAsp 380
 |||||
 Db 1081 TATAATGCCAGAGCTGATGAGCTCCGAGGGAATAATGCTTGCTGCATCTACTAGAT 1140
 QY 381 ArgSerGlnThrArgleuGlnIleValleuIleSerProGluLeuPheIleProValGlu 400
 |||||
 Db 1141 CGCTCCCAAGACTCGCTGCAAGATATGTATGATCTCACCTGAATATTTATTCACAGTAA 1200
 QY 401 AspAspValMetGluArgGlnArgGluIleGluSerValProAspSerValThrProLeu 420
 |||||
 Db 1201 GATGATGTTTATGGAAGAGCAAGAGACTCATTTGAGTCAATGCTGATTTGTGAGCCCACTA 1260
 QY 421 IleIleTyrgluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
 |||||
 Db 1261 ATATATCATATAGAAACAAACAGACATCTGATTAATTCATGATGACATCTTTCATGTTTTT 1320
 QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
 |||||
 Db 1321 CCCCAAGTCAACAGAGAAATTTGAGTTTATTTTTCCTGATGCAAAACAGGTTTC 1380

QY 461 ArgHisleuTyrrLysIleThrSerIleleuLysGluSerTyrrLysArgSerSerGly 480
 |||||
 Db 1381 GGTATTTATTAACAAATTAATTAATCTATTTTAAAGAAAGCAAAATTAACATCCAGTGT 1440
 QY 481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
 |||||
 Db 1441 GGGCTGCTGCTCCAGTGAATTTCAAGTGTCTATCAAAAGAGAGATGCAATTTACAGT 1500
 QY 501 GlyGluTrpGluValleuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
 |||||
 Db 1501 GGTGAATGGGAAGTTCCTTGCCCGCATGTGATCAATATCCAAAGTGTGTAAGTCAGAGG 1560
 QY 521 LeuValTyrrPheGluGlyThrLysAspSerProGluGluHisHisleuTyrrValIleSer 540
 |||||
 Db 1561 CTGGTATTTTAAAGCACCAAGACCTCCCTTTAGAGATCACTGTACATGATAGT 1620
 QY 541 TyrrValAsnProGlyGluValThrArgleuThrAspArgGlyTyrrSerHisSerCysCys 560
 |||||
 Db 1621 TACGTAAATCCTGAGAGGTGAGCAAGCTGACGAGCTGCTACTACATTCATCTTGCTGC 1680
 QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrrSerAsnGlnLysAsnProHisCys 580
 |||||
 Db 1681 ATCAGTCAGCACTGTGACTTCTTATTAAGTAAATAGTAACCAAGAAATCCACACTGT 1740
 QY 581 ValSerleuTyrrLysleuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 |||||
 Db 1741 GTGTCTCTTTACAGGCTATTAAGTCTGAAAGATGACCACCACTTGCAAAACAGAAATTT 1800
 QY 601 TrpAlaThrIleleuAspSerAlaGlyProleuProAspTyrrThrProProGluIlePhe 620
 |||||
 Db 1801 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCGTGACTATCTCCACAGAAATTTTC 1860
 QY 621 SerPheGluSerThrThrGlyPheThrleuTyrrGlyMetleuTyrrLysProHisAspLeu 640
 |||||
 Db 1861 TCTTTGAAAGTACTCTGATTTTACATGTGAGATGCTTCACAGCTTCATGATCTTA 1920
 QY 641 GlnProGlyLysAspTyrrProThrValleuPheIleTyrrGlyGluProGluValGlnLeu 660
 |||||
 Db 1921 CAGCTCGAAGAAATATCTCTACTGTGCTGTCTATATATGAGTGTCTCCAGGTGCACTTG 1980
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrrPheArgleuAsnThrleuAlaSerleuGly 680
 |||||
 Db 1981 GTGAATATATCGTTTAAAGAGTCAAGTATTTCCGCTTGAAATCCCTAGCCTCTAGGT 2040
 QY 681 TyrrValValValIleAspAsnArgGlySerCysHisArgGlyLysleuLysPheGluGly 700
 |||||
 Db 2041 TATGTGTTGTAGTATGATGACCAAGGGATCTGTGACCGAGGCTTAAATTTGAAGGC 2100
 QY 701 AlaPheLysTyrrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrr 720
 |||||
 Db 2101 GCCTTTAAATATTAATATGGGTCAAAATGAAATTTGACGATGAGTGGAAAGACTCCAAATAT 2160
 QY 721 LeuAlaSerArgTyrrAspPheIleAspLeuAspArgValGlylleHisGlyTrpSerTyrr 740
 |||||
 Db 2161 CTAGCTTCTCGATATGATTTTCATTGACTTATGATCGTGGGCGATCCACGAGCTGCTAT 2220
 QY 741 GlyGlyTyrrleuSerleuMetAlaMetGlnArgSerAspIlePheAspValAlaIle 760
 |||||
 Db 2221 GAGGATACCTCTCCCTGATGAGCATTAATGCAAGAGGTCAAGATCTTCCAGGGTTGCTAAT 2280
 QY 761 AlaGlyAlaProValThrleuTrpIlePheTyrrAspThrGlyTyrrThrGluArgTyrrMet 780
 |||||
 Db 2281 GCTGGGCCCCAGTCACTCTGTGATCTTTGATGATACAGATACACGAAACCTTATATG 2340
 QY 781 GlyHisProAspGlnAsnGluGlnGlyTyrrTyrrleuGlySerValAlaMetGlnAlaGlu 800
 |||||
 Db 2341 GGTCAACCTGACCAAGATGAACAGGCTATTACTTATAGATCTGTGGCATGCAAGCAGAA 2400
 QY 801 LysPheProSerGluProAsnArgleuLeuLeuHisGlyPheleuAspGluAsnVal 820
 |||||
 Db 2401 AAGTTCCCTCTGAAACCAATCGTTTACATGCTTACATGAGTTTCTGTGATGAAATGTC 2460
 QY 821 HisPheAlaHisThrSerIleleuLeuSerPheleuValArgAlaGlyLysProTyrrAsp 840

Db 2461 CATTTGCACATACAGATATATCTAGTGTTTTGTAGTGAAGGCTGGAAACCATATGAT 2520
Qy 841 leuglnlletyrproglngluarhiserlleargvalproglusegylglnhlsr 860
Db 2521 TTACAGATCTATCTCTCAGAGAGACAGCATTAAGATCTCTGATCGGAGAACATTAAT 2580
Qy 861 GluLeuHisLeuLeuHisTyrLeuGlnJuaenleugliserargilealaileuhs 880
Db 2581 GAATGCATCTTTTGCACTACCTTCAGAAAACCTTGATCAGTATTGCTCTTAAAA 2640
Qy 881 Valile 882
Db 2641 GTGATA 2646

RESULT 3
US-09-976-674-2
; Sequence 2, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akimsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2671
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-976-674-2

Alignment Scores:
Pred. No.: 0 Length: 2671
Score: 4700.00 Matches: 882
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-070-464-1 (1-882) x US-09-976-674-2 (1-2671)

Qy 1 MetAlaAlaIaIaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 8 ATGGCAGCAGCAATGGAACAGAACAGACCTGGGTCTTGAGATTTGAAACTGCCGACTGT 67
Qy 21 GlnGluAsnIleGluSerGlnAspArgProIysLeuGlnProPheTyrValGluArgTyr 40
Db 68 GAGGAGATATTTGATATACAGAGATCGGCTTAATTGAGCTTTTATGTGACCGGTAT 127
Qy 41 SetTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 128 TCCGCGGTGAGCTTAAGAAAGCTGCTTGCCGATACCGAAATATCATGCTACATATG 187
Qy 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 188 GCTTAAGCACAACATGATTTTCATGTTGTGAAAGGATATCCAGATGGACCTCATTTCA 247
Qy 81 AspArgIleTyrTyrLeuAlaMetSerGlyLysAsnArgLysLeuThrLeuPheTyrSer 100
Db 248 GACAGATCTATTAACCTTGCCATGCTCGTGAGAACAGAGAAATACACTGTTTATTTCT 307
Qy 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db 308 GAAATTCCTCAAACTAATCAATAGACGCACTCTTAATGCTCTCTTGAAAGCCCTTTTG 367
Qy 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGlnGluLeuLeuArg 140

Db 368 GATCTTTTTCAGGCAACACTGAGATATGATATCTGAGAGAAACATTTAAGA 427
Qy 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db 428 GAAAGAAAGCAATTTGGAACAGTCGGAATGCTTTCAACATATATACCAAGAGAGTGA 487
Qy 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspArgLysIleProGlnGly 180
Db 488 ACATTTCTGTTTCAAGCCGGTAGTGAAATTTATCATCTTAAGAAAGAGGCCCAAGGA 547
Qy 181 PheThrGlnGlnProLeuArgProAsnLeuValGlnThrSerCysProAsnIleArgMet 200
Db 548 TTTACGCAACACCTTTTAAGCCCAATCTAGTGAACCTAGTTCTCCACATACGATG 607
Qy 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db 608 GATCAAAATTTATGCCCTGCTGATCCAGACTGATGCTCTTTATATACATAGCAACGATATT 667
Qy 221 TrpIleSerAsnIleValThrArgGlnGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db 668 TGGATATCTTAACATCTGATACCAAGAGAAAGAGACTCATTAATGTCACAAATGACTTA 727
Qy 241 AlaAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
Db 728 GCCAACATGGAAGAAAGATGCCAGATACGCTGAGTCCCTACTTTGTTCTCAAGAAAGA 787
Qy 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysAlaGlnThrThrProSerGlyGly 280
Db 788 TTTGATAGATATCTGCTATTTGGTGTGTCCTCAAAAGCTGAACACATCCCAAGTGCTG 847
Qy 281 LysIleLeuArgIleLeuTyrGlnGluAsnAspGlnSerGlyValGluIleIleHisVal 300
Db 848 AAAATCTTAAGATCTATATGAAAGAAATGATGAATCTAAGTGAAATATTCATGTT 907
Qy 301 ThrSerProMetLeuGlnThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db 908 ACATCCCTATGTTGGAACAGAGAGGACGATTCATTCCTGTTATCTTAACACGGTACA 967
Qy 321 AlaAsnProLysValThrPheLysMetSerGlnIleMetIleAspAlaGlnGlyArgIle 340
Db 968 GCAAAATCTTAAGCATCTTTTAAGATGTCAGAAATATATGATTTGCTGAAGAGAGATC 1027
Qy 341 IleAspValIleAspLysGlnLeuIleGlnProPheGluIleLeuPheGlnGlyValGlu 360
Db 1028 ATGAGATGCTATGATTAAGAACTAATTCACCTTTTGAAGATTCATTTGAAAGGCTTGA 1087
Qy 361 TyrIleAlaArgAlaGlyTrpThrProGlnGlyLysTyrValaTrpSerIleLeuLeuAsp 380
Db 1088 TATATTGCCAGAGCTGATGATGACTCTGAGAGAAATATGCTTGGTCCATCTATAGAT 1147
Qy 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGlnLeuPheIleProValGlu 400
Db 1148 CGCTCCCAAGCTCGCTGAGATAGTGTGATCTCACTGATATATTATCCCAAGTGA 1207
Qy 401 AspAspValMetGluArgGlnArgLeuIleGlnSerValProAspSerValThrProLeu 420
Db 1208 GATGATGTTATGGAAGGCGAGACATTCATGAGTCACTGCTTGTGAGACGCCACTA 1267
Qy 421 IleIleTyrGlnGlnThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db 1268 ATTATCTATTAAGAAACACAGACATCTGATTAATTCATACATCTTTTCACTGTTTTT 1327
Qy 441 ProGlnSerHisGlnGlnGlnIleGluPheIlePheAlaSerGlnCysLysThrGlyPhe 460
Db 1328 CCCCAAGTCAGAAAGAGAAATTTGAGTTATTTTGGCTCTGAATGCAAAACAGGTTTC 1387
Qy 461 ArgHisLeuTyrTyrIleThrSerIleLeuLysGlnSerTyrTyrLysArgSerSerGly 480
Db 1388 CGTCAATTTATACAAATTTACATCTATTTTAAAGAAAGCAAAATTAACATCAGTGGT 1447
Qy 484 GlyLeuProAlaProSerAspPheLysCysProIleLysGlnGlnIleAlaIleThrSer 500

Db 1448 GGGCTGCTGCTCCAAGTATTTCAAGTGCTCCATCAAGAGAGATAGCAATTACAGT 1507
Qy 501 G1YGLTTPGLIValLeuGIYArGHISGIYSerANIIeGINValAepGIValArGArG 520
Db 1508 GGTGAATGGAAAGTCTTGCGCGGATGATCTAAATATCAAGTATGATGAATCGAAGG 1567
Qy 521 LeuValTYrPheGIuGIYThrLYsAAsPSePProLeuGIuHIeHIeLeuTYrValIser 540
Db 1568 CTGGATATATTTGMAAGCACCAAGACTCCCTTTAGACATCACTGAGCTAGTACGT 1627
Qy 541 TYrValAsnPProGIuValValThArGLeuthrAspArGIYTYrSerHISerCYsCYs 560
Db 1628 TAGCAAAATCTGGAGAGGTGACAAAGCTGACGACCGTGGCTACTACATTCTGCTCG 1687
Qy 561 IISerGINHISCYeAsPhePheIIeSerLYrSerAsnGINLYsAsnPProHISCYs 580
Db 1688 ATCAGTCAGACTGAGACTTCTTTATATAGTATAGTATACAGAGATATCACACTGT 1747
Qy 581 ValSerLeuTYrLYsLeuSerSerProGIuAsPAsPProThCYeLYrThrLYsGIuPhe 600
Db 1748 GTGTCTCTTACAGCTATCAAGTCTCTGAAGATGACCAACTTGCAAAACAAAGAAATT 1807
Qy 601 TTPAlArHrIIeLeuAsPSeRAlaGIYProLeuProAsPTYrThrProProGIuIIePhe 620
Db 1808 TGGGCAACCATTTGGATTGACAGAGTCTCTTCTGACTATACCTCCAGAAATTTTC 1867
Qy 621 SerPheGIuSerThrThGIYpHeThrLeuTYrGIYMeLeuTYrLYsProHISAsPLeu 640
Db 1868 TCTTTTGAAGTACTACGATTTTACATTGATGGATATCTCTACAGCTCATATATCTCA 1927
Qy 641 GINProGIYLYsLYsTYrProThrValLeuPheIIeTYrGIYGIYProGINValGIuLeu 660
Db 1928 CAGCTCGAAGAAATATCTCTACTGTGCTGTTCATATATGAGTGCTCTAGTGCACTTG 1987
Qy 661 ValAsnAsnArGpHeLYsGIYValLYsTYrPheArGLeAsnThrLeuIIaSerLeuGIY 680
Db 1988 GTGAATTAATCGATTTAAAGAGATCAAGTATTTCCCTGAATACCTCAAGCTCTAGGT 2047
Qy 681 TYrValValValValIIeAsPAsnArGpLYsSerCYsHISArGIYLeuLYsPheGIuGIY 700
Db 2048 TATGTGTGTGTAGTATGACAAACAGGGATCTCTGACCGAGGCTTAAATTTGAAGGC 2107
Qy 701 AlaPheLYsTYrLYsMeGIYGINIIeGINIIeAsPAsnGIYValGIuGIYLeuGINTYr 720
Db 2108 GCCTTAAATATTAATATGGGTCAAAATAGAAATTTGACGATCAGTGGAAAGACTCCAAAT 2167
Qy 721 LeuAlaSerArGIYTYrAsPheIIeAsPLeuAsPArGIYIIeHISGIYTYrPSeTYr 740
Db 2168 CTAGCTTCTCGATATGATATTCAATTGACTTACGTCGAGCATCCACGGCTGCTCTAT 2227
Qy 741 GIYGIYTYrLeuSerLeuMeRAlaLeuMeGIYArGSeRAlaPheArGIYAlaIIe 760
Db 2228 GAGAGATACCTCTCCCTGATGGCATTTAAAGCAGAGGTCAAGATATCTTCAGGGTTCAT 2287
Qy 761 AlaGIYAlaProValThrLeuThrPIIePheTYrAsPThrGIYTYrThrGIuArGIYrMeC 780
Db 2288 GCTGGGGGCCCAAGTCACTGTGGATCTTCTATGATACAGGATACAGGAAGTTATATG 2347
Qy 781 GIYHISProAsPGIYAsnGIuGINGIYTYrTYrLeuGIYSeRValIIaMeGIuAlaGIu 800
Db 2348 GGTCAACCCGACCAAGTAAACAGGGCTATTAAGTATCTGTGGCCATGCACAGAGAA 2407
Qy 801 LYsPheProSeRGIYProAsnArGLeLeuLeuLeuHISGIYPhLeuAsnGIuAsnVal 820
Db 2408 AAGTTCCTCTGAAACCAATATGTTTACTGCTCTTACATGATTTTCTGGATAGAAATGTC 2467
Qy 821 HIsPheAlaHISerThSerIIeLeuLeuSerPheLeuValArGIYLYsProTYrAsP 840
Db 2468 CATTTTGACATATACAGTATATTAAGTATTTTATAGTGAAGGCTGGAAGCCATATGAT 2527
Qy 841 LeuGINIIeTYrProGINuArGIYHISerIIeArGIYValProGIuSerGIYGIuHISTYr 860
Db 2528 TTACAGATCTATCTCAAGAGAGACACAGCATTAAGATTCGAAATCGGAGAAACATTA 2587

Qy 861 GIuLeuHISLeuLeuHISerTYrLeuGINuAsnLeuGINLYsSerArGIYAlaIIaIIeLYs 880
Db 2588 GAACATGATCTTTGACATCACTTCAGAAACCTTGATCAAGTATGCTGCTCAAAA 2647
Qy 881 ValIIe 882
Db 2648 GTGATA 2653
RESULT 4
US-10-170-789-37
; Sequence 37, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
; TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USBS THEREOF
; FILE REFERENCE: 10448-191001
; CURRENT APPLICATION NUMBER: US/10/170,789
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797,039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882,166
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,078
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/934,406
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26052
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06

PRIOR APPLICATION NUMBER: PCT/US01/07074
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 60/187,420
PRIOR FILING DATE: 2000-03-07
NUMBER OF SEQ ID NOS: 63
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 37
LENGTH: 3143
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (229) ... (2874)
US-10-170-789-37

Alignment Scores:

Pred. No.:	0	Length:	3143
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12	Indels:	0
DB:		Gaps:	0

US-10-070-464-1 (1-882) x US-10-170-789-37 (1-3143)

QY 1 MetAlaIaIaIaMetGluThrGluGluGluGluValAlaThrPheGluThrAlaAspCys 20
DB 229 ATGGACACAGCAATGAGAAACAGAACAGCTGGGTGTGAGATTGAAACGCGACTGT 288
QY 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
DB 289 GAGAGAAATTGATTCACAGGATCGGCTTAATTGAGCCTTTTATGTTGAGCGGTAT 348
QY 41 SerTyrSerGluLeuLysLysLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
DB 349 TCTTGAGTCACTTAAAGCTGCTTGCATCCAGATCCAGAAATATCATGGCTACATGATG 408
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
DB 409 GCTTAGGCACACATGATTCATGTGTTGTGAGAGATGATCCAGATGACCTCATTTCA 468
QY 81 AspArgLysTyrTyrLeuAlaMetSerGlyGluAsnArgLysAsnThrLeuPheTyrSer 100
DB 469 GACGAGATCATTAACCTTGCATGTCTGTGAGAAACAGAAATATACACTGTTTATCT 528
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTyrProLeuLeu 120
DB 529 GAAATTCCTCAAACTATCATATAGACAGCACTTAATGCTCTGTTGAAAGCTCTTTTG 588
QY 121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
DB 589 GATCTTTTCAAGCACACAGTGAAGTATGATGATCTCGAAGAGAACTATTAGA 648
QY 141 GluArgLysArgLysGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB 649 GAAAGAAACGCTTGGACAGCTGGAATGCTCTTACATATATCCAGAAAGTGA 708
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
DB 709 ACATTTCTGTTTCAAGCCGCTAGTGAATTAATCATCACTTAAGATGAGGCGCAACAGA 768
QY 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB 769 TTATACGAAACCTTTAAGGCCCAATCTAGTGAAACTAGTGTCCCAACATACGATG 828
QY 201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle 220
DB 829 GATCCAAATTAATGACCTGCTGATCCAGACTGAGTTGCTTTTATACATACCAACGATAT 888
QY 221 TrpIleSerAsnIleValThrArgGluGluArgLeuThrTyrValHisAsnGluLeu 240
DB 889 TGGATATCTTAACATCGTAACAGAAAGAAAGAGACTCATATATGCAATATGAGCTA 948

QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 260
DB 949 GCCAACATGGAAGAGATGCCAGATCAGCTGAGTCCCTACTTGTCTTCCAAAGACA 1008
QY 261 PheAspArgTyrSerGlyTyrTyrProCysProLysAlaGluThrThrProSerGlyGly 280
DB 1069 TTTCATAGATATTCCTGCTATTTGGTGTGTCCAAAAGCTGAAACATCCAGCTGCTGT 1068
QY 281 LysIleLeuArgLysLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB 1069 AAAATTCCTGAATCTATATGAAAGAAATGATCAATCTGAGGAGAAATTAATCATGTT 1128
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB 1129 ACATCCCTATGTTGAAACAGAGGCGCAGATTCATTCGTTATCTTAAGACGATCA 1188
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB 1189 GCAATCTTAAAGTCACTTTTAAAGATGTCAAGAAATATGATGATGCTGAAGAGATC 1248
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB 1249 ATGATGTCAATAGTAAAGAACTAATCAACTTTTGAGATTCATATTTGAAAGAGTTGAA 1308
QY 361 TyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIleLeuLeuAsp 380
DB 1309 TATATTCACAGAGCTGATGATGACTCTGAGAGAAATATGCTGTGCTCATCTACTGAT 1368
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB 1369 CGCTCCAGACTGCGCTGCGAGATAGTGTGATCTCACTGAATATTAATCCAGTGA 1428
QY 401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
DB 1429 GATGATGTTTGAAGAGCAGACATCATTAAGTCACTGAGTGTCTGAGCCGCACTA 1488
QY 421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
DB 1489 ATTAATCTATAGAAACACAGACATCTGGATTAATTCATGACATCTTTCATGTTTT 1548
QY 441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
DB 1549 CCCCAAGTCAACAGAGGAAATTTGAGTTATTTTGGCTCTGAATCCAAACAGTTTC 1608
QY 469 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
DB 1609 CGTCATTTATACAAATTAATCATCATTTTAAAGAAACAAATATAACATCCAGTGT 1668
QY 481 GlyLeuProLysProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB 1669 GGGCTGCTGCTCCAAGTGAATTCAGATGTCCTATCAAAAGAGATGACATTTACAGT 1728
QY 501 GlyIleThrGluValLeuGluArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB 1729 GGTGAATGGGAATTTCTTGGCCGCGCATGTGATCTTAATATCCAGTTGATGAAGTCAGAGG 1788
QY 521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValIleSer 540
DB 1789 CTGGTATATTTGAAGCACCAAGACTCCCTTTAGAGCATCACCTGATCGTAGTACGT 1848
QY 541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB 1849 TACGTAAATCTTGAAGAGGAGCAAGCTGACCTGAGCTTACCTCACATTTCTGCTGC 1908
QY 561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
DB 1909 ATCAGTCAGACATGTAATCTTTATAGTATAGTATAGTAAACAGAAATCCACACTGT 1968
QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
DB 1969 GTGTCCCTTTACAGACTATCAAGTCTCTGAAGATGACCAACTTGCAAAACAAAGATTT 2028
QY 601 TrpAlaIleThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620

Db 2029 TGGGGCCACCATTTGGATTCCAGCAGGTCCTCTCTGACTATACCTCCCGAAATTTTC 2088
Qy 621 SerPheGluSerThrThrGlyPheThrLeuTyrglyMetLeuTyrlsPheProHisAspLeu 640
Db 2089 TCTTTGAAAGTACTACTGATTTTACATTTGATGAGATGCTCAACAGCCTCATGATCTA 2148
Qy 641 GlnProGlyLysLysTyrrProThrValLeuPheIleTyrglyGlyProGlnValGlnLeu 660
Db 2149 CAGCTGGAAAGAAATATCTACTGCTGCTCATATATGGGCTCTCAGGTGCGATG 2208
Qy 661 ValAsnAsnArgPheLysGlyValLysTyrrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db 2209 GTGATTAATCGGTTTAAAGAGCTCAAGTATTTCCCTTGAAATACCTTAGCCTCTAGGT 2268
Qy 681 TyrValValValValLysAspAsnArgLysSerCysHisArgGlyLeuLysPheGluGly 700
Db 2269 TATGCTGTGATGATAGACAAACAGGGGATCTCTGACCGAGGGCTTAAATTTGAAAGGC 2328
Qy 701 AlaPheLysTyrrLysMetGlyGlnIleGlnIleAspAspGlnValGluGlyLeuGlnTyrr 720
Db 2329 GCCTTTAAATATATAATGGGTCAATAGAAATGACAGATCAGGTGGAAAGACTCCAAATAT 2388
Qy 721 LeuAlaSerArgTyrrAspPheIleAspLeuAspArgValGlyLysGlyTrrPserTyrr 740
Db 2389 CTAGCTTCTCGATATGATTTCTTCACTTAGATCGGTGGGATCCAGGCTGCTCTAT 2448
Qy 741 GlyGlyTyrrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2449 GGAGGATGCTCTCCCTGATGCGATTAATGACAGAGGTCAAGATATCTTCAGGGTTGCTAT 2508
Qy 761 AlaGlyAlaProValThrLeuTrpIlePheTyrrAspTrpGlyTyrrThrGluArgTyrrMet 780
Db 2509 GCTGGAGGCCCAAGTACTCTGGATCTTCTTATGATACGAGTACACGGAACGTTATATG 2568
Qy 781 GlyHisProAspGlnAsnGluGlnGlyTyrrTyrrLeuGlySerValAlaMetGlnAlaGlu 800
Db 2569 GGTGACCCCTGACCAAGATGAACAGGGGCTATTACTTAGAGTCTGTGGCCATGGAAACAGAA 2628
Qy 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db 2629 AAGTCCCCCTCGAACCAAAATGTTACTGCTTACATGATGTTCTCTGATGAGATGTC 2688
Qy 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTrpArg 840
Db 2689 CATTTTGACATACACAGATATATTACTGACTTTTATAGTAGGGCTGGAAAGCCATATGAT 2748
Qy 841 LeuGlnIleTyrrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrr 860
Db 2749 TTACAGATCTATCTCCAGAGAGACACAGCATPAAGATTCTCGAATCCGGAGAACTATAT 2808
Qy 861 GluLeuHisLeuLeuHisTyrrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db 2809 GAACGCAATCTTTTGACATACCTTCAGAAACCTTGATGACGATATGCTGCTTAAAA 2868
Qy 881 ValIle 882
Db 2869 GTGAT 2874

US-09-976-674-12
; Sequence 12, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117

US-09-976-674-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-12

Alignment Scores:
Pred. No.: 0 Length: 4829
Score: 4680.00 Matches: 882
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.57% Indels: 2
DB: 10 Gaps: 0

US-10-070-464-1 (1-882) x US-09-976-674-12 (1-4829)

Qy 1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGlnIlePheGluThrAlaAspCys 20
Db 214 ATGGCAGCAGCAATGAAAGAAACAGACGGGTGTGAGATATTTGAAACTGGGACTGT 273
Qy 21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrrValGluArgTyrr 40
Db 274 GAGGAGAAATATTGAATACAGGATCGGCTAAATTTGAGCCTTTTATGTTGAGCGGAT 333
Qy 41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrrHisGlyTyrrMetMet 60
Db 334 TCTGGAGTCAAGCTTAAAAAACCTGCTTGCATGCCAGAAATATCATGGCTACATGATG 393
Qy 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProHisSer 80
Db 394 CTTAAAGCACAACATGATTTATGTTGTGAAGAGAAATGATCCAGATGAGCTTCATTTCA 453
Qy 81 AspArgIleTyrrTyrrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrrSer 100
Db 454 GACGGAATCTATTACCTTGCCATGTCTGTGAGAACAGAGAAATACACTGTTTATTCT 513
Qy 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db 516 GAAATTCACCAAAACATCATATAGACAGCAGCTTAATGCTCTCTGGAAACCTCTTTTG 573
Qy 121 AspLeuPheGlnAlaThrLeuAspTyrrGlyMetTyrrAspArgGluGluGlnLeuArg 140
Db 574 GATCTTTTTCAGGCAACCTGACCTATGAAATGATATCTTCAGAAAGAACTATTAAGA 633
Qy 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrrAspTyrrHisGlnIleSerGly 160
Db 634 GAAAGAAACCATTTGGAACAGTGGAAATTCCTTTACGATATACACCAAGAAAGTGA 693
Qy 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrrHisValLysAspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTTCAAGCCGGTAGTGAATTTATACGATTAAGATGGAGGCCCAAGAA 753
Qy 181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db 754 TTACGCAACCACTTTTAGGCGCAATCTAGTGAATACTATGTTCCCAACATACGAGATG 813
Qy 201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db 814 GATCCAAATATATGCGCTCGATCCAGACCTGGAATGCTTTTATATACATAGCAAGATAT 873
Qy 221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrrValHisAsnGluLeu 240
Db 874 TGGATATCTTAACATCGTAACAGAGAAAGAGAGACTCACTTATGTGCACAAATAGAGCTA 933
Qy 241 AlaAsnMetGluGluAspAlaArgSerArgIleGlyValAlaThrPheValLeuGlnGluGly 260
Db 934 GCCAATCATGAAAGAGATGCAATCAGCTGAGTGGAGTCTTGTCTTCCAGAAAGAA 993
Qy 261 PheAspArgTyrrSerGlyTyrrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280

Db	994	TTTGATAGATATCTCGGCTATTCGGTGGTGTGTCCAAAGCTGAACAACTCCAGTGGTGGT	1053
Qy	281	LYSIIIELEUARGIIIEUITYGIUUBNAAPGJUSERGIVUAGIUIEIIIEIIEVAL	300
Db	1054	AAAAATCTTAGAATTCTATCTATGAAAGAAATGATGATCTGAGGTGGAATTAATTCATGTT	1113
Qy	301	THISERRPROMETLEUGIUTHIRKGRKRLAAPSERRPHEATGTYRPROLYSTHGIYTHR	320
Db	1114	ACATCCCCCATGTGTGAAACAAAGAGGGCAGAGTTATATTCCTGTATCCCTAAACAGGTACA	1173
Qy	321	ALAENPROLYSEVALTHIRHELYMESESGUIIEMETIEAEPALIGIUIYARGIIIE	340
Db	1174	GCAATTCCTAAAGTACTCTTTAAGATGTCCAGAAATATGATGTGATGCTGGAAGGAAGATC	1233
Qy	341	IIAEPVALIIIEAPSYGSIUENIIEGNIIPROPHGIUIIEUUPHEGIUIYVAGIUB	360
Db	1234	ATGATGTCATGATGATGAAGAACTAATTCAACTTTGAGATTCTATTGGAAGAGTTGAA	1293
Qy	361	TYRIIEAIIAARGALAGIYTRPHRPROGIUGIUYSTYRAIATPSEIIIELEUENAP	380
Db	1294	TATATTTGCCAGAGCTGAGATGAGCTCTTACAGGAAATATGCTGTGATCTCATCTCATGAT	1353
Qy	381	ARGSERGINTHIRXLEUGIUIIEVALLEUIESERPROGIULEPHIEIIEPROVALGIUB	400
Db	1354	CCCTCCCAAGCTCCGCTTACAGATAGTGTGATCTCACTGAAATTTATTCCTCCAGTACAA	1413
Qy	401	ASPAAPVALMEGIUARGIUNARGLEUIIEGUSEVALIIPROAPSERVALTHIRPROLEUB	420
Db	1414	GATGATGTTATAGGAAGGACAGACTCATTTGAGTCAGTGGCTGATTCGTGACGCCACTA	1473
Qy	421	IIIEIETRYGIUGIUTHIRASPIIETRPIIEANIIIEHISAPRIIEPHIEIIVALPHE	440
Db	1474	ATTATCTTGTGAAGAAACACAGACACTCGGATTAATATCCATGTGATCTTTCAATGTTT	1533
Qy	441	PROGINSERHISGIUGIUGIUIIEGIUBHEIIEPHIEIASERGIUBCYLYSTHGIYIPHE	460
Db	1534	CCCCAAGTCCAGAGAGAGAAATGAGTTATTTTGCTCTGATGTCGAATCCAAACAGGTTTC	1593
Qy	461	ARGHISLEUITYLYSIIETHSERIIIELEUBYSJUSELITYRYLYSARGSERGIY	480
Db	1594	CGTCATTTATACAAATTAATTCATCTATTTTTAAAGGAACAAATTAACGATCCAGTGGT	1653
Qy	481	GIYLEUPROALIPROSERASPHELYCYSPROIIELYSGIUGIUIIEALIIETHSER	500
Db	1654	GGCGCTGCTGCCAAGATTTCAAGTGTCTTATCAAAAGAGGATAGCAATTAACAGT	1713
Qy	501	GIYGIUTPRGIUVALLEUGIYARHISGIIYSERANIIIEGINVALAEPGIUVALARGX	520
Db	1714	GGTGATGGAGAGTTCTTGCGCGGATGATCTAATATCCAAAGTGAAGAGTACGAAGG	1773
Qy	521	LEUVALIYRPHGIUGIYTHIRLYSAPSERPROLEUGIUNISHISLEUYRVALIASER	540
Db	1774	CTGGATATTTTGAAGGACCAAGACCTCCCTTTAAGGATCACTGTAGTACGTACGT	1833
Qy	541	TYRVALANPPOGIUGIUBVALTHIRARGLEUTHIRAPARGIUYRSEHISESCYSCYS	560
Db	1834	TACGTAATTCCTGGAGAGGTACAAAGCTGACTGACCGTGGCTACTCATCTTTGCTGC	1893
Qy	561	IIIESERGINHISCYASAPRPHHEIIEERTYSTYRSEARANGIUIEAPNPROHISCYS	580
Db	1894	ATCAGTCCAGCACTGACTCTTTATATAGTAAGTATACCAAGAAATCCACACTGT	1953
Qy	581	VALSERLEUYRILYLSERSESERPROGIUBAPAPPROTHIRCYLSYRTHIRLYSGIUBPHE	600
Db	1954	GTGTCCTCTTACMAAGCTATCAAGTCTGAAATGACCACTTGCAAAACAAAGAAATTT	2013
Qy	601	TRPALATHRIIELEUAPSERALAGIYPROLEUBPROAPRYRTHIRPROGIUIIIEPHE	620
Db	2014	TGGGCAACCATTTGGATTTCAGCAGGTCTCTTCTGACTATCTCTCCCGAATTTTC	2073
Qy	621	SERPHEGIUSERTHIRTHGIYPHETHIRLEUITYGIIEMETLEUITYRPIROHISAPRLEUB	640
Db	2074	TCITTTGAAATGACTACGTGATTTACATTTGATAGTGAGTGTCTTACAAAGCCCTCAAGATCTA	2133

OY		641	GINPPOGIVLYVelytYrProothrtAlLeubhellEtyrglygyl--ProginvAIGlnl	660
Dd		2134	CAGCCTGGAAAGAAAATATCTACGTGTGCCTTCATATAATGGTGGCTCCCTCAGGTGCACT	219
OY		660	eUValAsnABsnArghPhelYsgIyValLySTyrPheArgLeuasnThrLeuAlaSerLeuc	680
Dd		2194	TGGTGAAATATACGGTTATTAAAGAGCAAGATATTTCCCGCTGAATACCCTAGCCTCTTAG	225
OY		680	IYTYrVaIvalValIalIleapanaargIyserCyshisArgGilyLeuLysPheJug	700
Dd		2254	GTTATGTGGTGTGTACTGTATGACAACAGGGGATCCTGTCAACGAGGGCTTAATTTGMAAG	231
OY		700	IYAlaPhelyTyrltYsmetGlygnIleGIuIIeasApogInvalGluJugLeuGINt	720
Dd		2314	GGCCTTTAAATATAAAATGGGTCAAATAGAATTTGAAGATCACGGTGAAGGCCATCCAAT	237
OY		720	YrLeuAlaserArgTyraSPheIleasPLeuasPAryValGIyIlehisGIYTTPserT	740
Dd		2374	ATCTAGCTTCTCGAATATATCTCATGTGACTTAGACGCGTGGGACTCCAGCGCTGGTCTT	243
OY		740	YRGlyGIYTYrLeuSerLeuMetAlaleuMetGlnarSergSerAspIleebeargValAlai	760
Dd		2434	ATGGAGGTACTCTCTCCCTGATGGCATTAATGACAGAGGTCAAGATATCTTCAGGGTTGCTA	249
OY		760	IleaGIYAlaPrOvalThriturprIlePhetyrasPTThngIYTYrThrgIuaqTYrm	780
Dd		2494	TTGCTGGGGCCCCCACTCACTCTGTGGATCTTCTMGATACAGGAATACCGMAACCTTATA	255
OY		780	etGIYHisPRoAspDlnanGluGlnGIYTYrTYrLeuGIYSerValAlametGlnAlag	800
Dd		2554	TGGGCACCTCGACCAAGATTAACAGGGCTATTACTTAGATCTGTGGCCATGCMAAGCAG	261
OY		800	IuLYrPhePRoserGIuPrOasnArgLeuLeuLeuHiseGIYPheLeuaspGIuasnv	820
Dd		2614	AAAAGTTCCTCCCTGMAACCAATCGTTACTGCTCTTACATGGTGTCTTCGATGAGAATG	267
OY		820	aLIsePhAlaHIethserIleleuLeuSerPheLeuValArgAlaalyLysProtYrA	840
Dd		2674	TCCATTTTGCACATACACGATATTACTGAGATTTTTTAAAGAGGCTGGAAAGCCATATGG	273
OY		840	sPLEuGlnIleTYrPROgIngluarGrHisSerIleArgValIProgIusergIyGuhist	860
Dd		2734	ATTTAACATCTATTCCTCAGAGACACACACATTAAGGTTCTCGAATCGGAGAACATT	279
OY		860	YrGIuLeuHileuLeuHietYrLeuGlnIuasenLeuGIYSerArgIleAlaalaleuL	880
Dd		2794	ATGAATGCACATCTTTTGACATACTTCAGAAACCTTGATCAGATTTGTCTGCTCTAA	285
OY		880	ySvalIle 882	
Dd		2854	AAGTGATTA 2861	
 RESULT 6 US-09-976-674-22 ; Sequence 2g Application US/09976674 ; Patent No. US20020115843A1 ; GENERAL INFORMATION: ; APPLICANT: Qi, Steve ; APPLICANT: Akinsanya, Karen ; APPLICANT: Riviere, Pierre ; APPLICANT: Junien, Jean-Louis ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV ; FILE REFERENCE: 70669 ; CURRENT APPLICATION NUMBER: US/09/976, 674 ; CURRENT FILING DATE: 2001-10-12 ; PRIOR APPLICATION NUMBER: US 60/240, 117 ; PRIOR FILING DATE: 2000-10-12 ; NUMBER OF SEQ ID NOS: 61 ; SOFTWARE: PatentIn version 3.1 ; SEQ ID NO 22 ; LENGTH: 4685 ; TYPE: DNA				

ORGANISM: Homo sapiens
US-09-976-674-22

Alignment Scores:

Pred. No.:	0	Length:	4685
Score:	4385.50	Matches:	834
Percent Similarity:	94.56%	Conservative:	0
Best Local Similarity:	94.56%	Mismatches:	1
Query Match:	93.31%	Indels:	48
DB:	10	Gaps:	1

US-10-070-464-1 (1-882) x US-09-976-674-22 (1-4685)

```

OY      1 MetAlaAlaIaMeCgluThrgluGluGluIlePheGluThrAlaAspCys 20
DB      214 ATGGGAGGAGCATGGAAACAGACAGCGGGGTGGAGATATTGAACTCGGACTGT 273
OY      21 GluGluAsnIleGluSerGluAspArgProLysLeuGluProPheTyrValGluArgTyr 40
DB      274 GAGGAGAAATATGAAATCAGACAGATCGGCTAAATTGGAGCCCTTTTATGTGAGCGGTAT 333
OY      41 SerTPSerGluLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetC 60
DB      334 TCTTGAGTCAAGCTTAAAGAGCTGTGGCGGATACAGAAATATCATGGCTTACATGATG 393
OY      61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
DB      394 GCTAAGGACCAACATGATTTGATTTGTGGAAGAGAAATGATCCAGATGGACCTCATTTCA 453
OY      81 AspaArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
DB      454 GACAGATCTATTAATCCTTCATGCTGTGGAGAAACAGAAATATCATGCTTTTATTTCT 513
OY      101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
DB      514 GAAATTCCTCAAAACATTCATAGAGACAGACAGCTTAAATCTCTCTTGGAAGCCCTCTTTG 573
OY      121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
DB      574 GATCTTTTCAAGCAACACTGGACCTATGGAATGTATTTCTCGAAGAAAGAACTATTAAAGA 633
OY      141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
DB      634 GAAAGAAACGATTTGGAACAGTGGAAATGCTCTTACGATTATCACCAAGAAAGTGA 693
OY      161 ThrPheLeuPheGluAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
DB      694 ACATTTCTGTTCAAGCCGGTAGTGAAATTTATCACGTTAAAGATGGAAGGCCACAAAGGA 753
OY      181 PheThrglnGluProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
DB      754 TTACGCAACAACTTTAAAGCCCAATCTAGTGAATAGTTGTCCCAACATACGGATG 813
OY      201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
DB      814 GATCCAAATTAATGCTGCTGATCCAGCTGGAATGCTTTTATCATGTGCAACCATATTT 873
OY      221 TrpIleSerAsnIleValThrArgGluGluArgGlyLeuThrTyrValHisAsnGluLeu 240
DB      874 TGGATATCTTAACATGCTAACCGAAGAAAGAAAGAAAGATCATCTTAATGTGCACATAGCTCA 933
OY      241 AlaAsnMetGluLysAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
DB      934 GCCAACATGGAAGAGATGCCAGATCAGCTGAGTGCCTACTTGTCTCCAAAGAAAGA 993
OY      261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
DB      994 TTTCATAGATATTCGGCTATTTGGTGTGTCCTCAAAAGCTGAACCACTCCCGTGTGTGT 1053
OY      281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
DB      1054 AAAATTTCTAGAAATCTATATGAAAGAAATGATGATGAGGTGGAATTTATCATGTT 1113

```

```

OY      301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
DB      1114 ACATCCCTATGTTGGAAACAGAGAGGCGAGTTATTCCTGTTATCTTAAACAGGTACA 1173
OY      321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
DB      1174 GCAAATTCCTAAAGTCACTTTTAAAGATGTACAGAAATTAAGATGATGATGCAAGAAAGATTC 1233
OY      341 IleAspValIleAspGlyGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
DB      1234 ATGATGTCATATGATTAAGAACTAAATCAACTTTTGAGATTTCAATTTGAAAGAGTTGAA 1293
OY      361 TyrIleAlaArgAlaGlyTyrTrpProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
DB      1294 TATATTCGAGAGCTCGATGAGCTCTCAGAGGAAATATGCTGTGCTCATCTACTAGAT 1353
OY      381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
DB      1354 CGCTCCAGACTCGCTCAGATAGTGTGATCTCACTGAATTTATTTATCCAGTAGAA 1413
OY      401 AspaAspValMetGluArgGluAlaArgLeuIleGluSerValProAspSerValThrProLeu 420
DB      1414 GATGATGTTATGGAAGGACAGACTCATGATGATGAGCTGATCTGTGTGACGCCACATA 1473
OY      421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
DB      1474 ATTATCTATGAAGAAACAAACAGACATCTGATTAATATCCATGACATCTTTTATGTTT 1533
OY      441 ProGlnSerHisGlyGluGluIleGluPheIlePheAlaSerGlyCysLysThrGlyPhe 460
DB      1534 CCCCAAGTCAAGAAAGAAATTAAGATTTATTTTGTGCTGATTCGAAATGCAAAACAGATTTC 1593
OY      461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerGly 480
DB      1594 GCTATTTATCAAAATTAATCAATTTTAAAGGAAACCAATATTAACAGATCCAGTGT 1653
OY      481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
DB      1654 GGGCGCTGCTGCCAAGGATTTTCAAGTCTCTATCAAAAGAGAGATACCAATTTCCAGT 1713
OY      501 GlyIleTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
DB      1714 GGTGAATGGAAGATTTCTTGGCCGCGATGATCAATATCAAAATGATGAAGTCAAGAAAG 1773
OY      521 LeuValTyrPheGluGluThrLysAspSerProLeuGlnHisIleLeuTyrValValSer 540
DB      1774 CTGATATATTTTGAAGGCACCAAGACTCCCTTTAAGACATCACCTGATGATGTCAGT 1833
OY      541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
DB      1834 TACGTAAATCTCGAAGAGGTACAAAGCTGACTGACCGTGGCTACTCATTCTTGCTGC 1893
OY      561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerArgGlnLysAsnProHisCys 580
DB      1894 ATCAGTCAAGACTGAGACTCTTTATTAAGTAAATAGTAAACCAAGAAATCCACACTGT 1953
OY      581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysGlySerThrLysGluPhe 600
DB      1954 GTGTCCCTTTCAAGACTATCAAGTCCGAAATATGACCAACTTCCAAACAAAGAAATTT 2013
OY      601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspArgTyrThrProProGluIlePhe 620
DB      2014 TGGGCAACCATTTTGGATTCAGT----- 2036
OY      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
DB      2036 ----- 2036
OY      641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
DB      2037 -----CCTCAGATCCAGTTG 2051
OY      661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

```

Accession	Gene	Protein	Length	Score
Db	2052	GGGATTAACGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTAGCTTACGTT	2111	
Qy	681	TYRVALVALVAL11LEAPSAANRGISERYSHISARGIYLEULYPHEGLUGLY	700	
Db	2112	TATGCTGTGTAGTGATAGACAAACAGGGATCTCTGCACCGAGGCTTAAATTTGAAAGC	2171	
Qy	701	ALAHELYETYLEMELGLYGLN1LEGLUL1EAPAPRG1NVALGLUGLYLEUGLNTYR	720	
Db	2172	GCCTTTAAATAAATAAATGGGTCAATATGAAATGTGACATCAGGTGGAAGGACTCCAAAT	2231	
Qy	721	LEUALASERARGTYRASPHE1LEASPLEUASPAVVALGL1LEHISGLYTRSERTYR	740	
Db	2232	CTAGCTTCTCGATATGATTTCTTTGACTTACATCGTGGGCAATCAGCGCTGGCTCAT	2291	
Qy	741	GLYGLYTYRLEUSERLEUMELALEUMELG1NARGSERASPL1EPHEARGVALAILE	760	
Db	2292	GGAGATACCTCTCCCTGATGGCAATTAATGACAGAGTCAGATATCTTCAAGGTTCTAT	2351	
Qy	761	ALAGLYALAPROVALTHLEUTR1LEPHERYRABPTHRYTYRTHGLUARGTYRMEC	780	
Db	2352	GCTGGGGCCCCAGTCACTCTGGATCTTCTATGTATACAGATACACGGAAGCTTATATG	2411	
Qy	781	GLYHISPROASRG1NANGLUG1NGLYTYRTRYLEUG1SERVAL1AMEGLALAGLU	800	
Db	2412	GGTCACTCTGACCAAGATGACAGGGCTTATCTTAGATCTGTGGCCATGACAGACAGA	2471	
Qy	801	LYSPHEPROSEGLUPROASNARGLEULEULEUHLISGLYPHELEUASPG1UANVAL	820	
Db	2472	AAGTTCCCTCGAACCAAAATGTTTACGTCTTACATGGTTTCTTGAGAGAAATGTC	2531	
Qy	821	HISPLEALAHSTHSER1LEULEUSERPHELEUVALAAGALGLYLRPTQTYRABP	840	
Db	2532	CATTTTGCATACATCACATATATTTACTGTAGTTTTTTAGTGAAGGGCTGGAACCATATGAT	2591	
Qy	841	LEUGN1LEYRPRG1NG1NARGHISSER1LEARGVAL1PROGLUSERGLYGLNHSERYR	860	
Db	2592	TTACAGATCTATCTCAGAGAGACACAGCATMAAGTTCTCGAATGGGGAACATTTAT	2651	
Qy	861	GLUEUHLISLEUHLIS1TYRLEUG1N1UANLEUG1SERARG1LEALALEULYS	880	
Db	2652	GAACTGCATCTTTTGCACCTTCAABAAAACCTTGATACAGATTTGCTGCTTAAAA	2711	
Qy	881	VAL11LE 882		
Db	2712	GTGATA 2717		
RESULT 7				
US-09-976-674-20				
; Sequence 20, Application US/09976674				
; Patent No. US20020115843A1				
; GENERAL INFORMATION:				
; APPLICANT: Qi, Steve				
; APPLICANT: Akinsanya, Karen				
; APPLICANT: Riviere, Pierre				
; APPLICANT: Junien, Jean-Louis				
; FILE REFERENCE: 70669				
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV				
; CURRENT APPLICATION NUMBER: US/09/976,674				
; PRIOR FILING DATE: 2001-10-12				
; PRIOR APPLICATION NUMBER: US 60/240,117				
; PRIOR FILING DATE: 2000-10-12				
; NUMBER OF SEQ ID NOS: 61				
; SOFTWARE: PatentIn version 3.1				
; SEQ ID NO 20				
; LENGTH: 4676				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
US-09-976-674-20				
Alignment Scores:				
Pred. No.: 0				
Score: 4385.00				
Length: 4676				
Matches: 831				

Percent Similarity:	94.22%	Conservative:	0
Best Local Similarity:	94.22%	Mismatches:	1
Query Match:	93.30%	Indels:	51
DB:	10	Gaps:	1
US-10-070-464-1 (1-882) x US-09-976-674-20 (1-4676)			
QY	1 MetAlaAlaMeGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys	20	
DB	214 ATGGCAGCGACGAATGAAACAGAACAGCTGGCTGTGAATATTGAACTGGCGACTGT	273	
QY	21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluGlyTyr	40	
DB	274 GAGGAGAAATATTGAATCAACGATGGCGCTTAATTGGACCTTTTATGTTGACCGCTAT	333	
QY	41 SerTPSerGlnLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet	60	
DB	334 TCCTGAGTCACTTTAAAGCTGCTTCCGATACAGAAATATCATGTGCTTCATGATG	393	
QY	61 AlaLysAlaProHisAspPheMetCysValLysArgAsnAspProAspGlyProHisSer	80	
DB	394 GCTAAGCACCACTGATTTCTGTTGGAGAGAAATGATCCAGATGACTCATCA	453	
QY	81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer	100	
DB	454 GACAAATCTTATTACCTTGCCATGCTGGTGAAACAGAAATAACATGTTTATATCT	513	
QY	101 GluIleProLysThrLisAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeuLeu	120	
DB	514 GAAATTTCCCAAACTATCATATGACGACAGCTTTAATGCTCTCTTGGAAGCTCTTTTG	573	
QY	121 AspleuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg	140	
DB	574 GATCTTTTTCAGGCAACACTGGACTATGAAATGATTTCTCGAAGAAAGAACTATTAGA	633	
QY	141 GluArgLysArgIleGlyTyrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly	160	
DB	634 GAAAGAAACCGATTGGAACAGTCGGAATGCTTCTTCGATTATCACCAAGGAATGGA	693	
QY	161 ThrPheLeuPheGlnAlaGlySerGlyLysTyrHisValLysAsnGlyGlyProGlnGly	180	
DB	694 ACATTTCTGTTTCAGCCGCGTATGTAATTTATTCACGTAAAGATGAGGGCCACAGGA	753	
QY	181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet	200	
DB	754 TTTAGCGAACAACCTTTAAGGCCAATCTAGTGAATAGTGTGCCAACATACGATG	813	
QY	201 AspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAsnAspIle	220	
DB	814 GATCCAAATATTATGCCCTGATCCAGATCCAGATTCGTTTATTCATAGCAACATATT	873	
QY	221 TrpIleSerAsnIleValThrArgGluGluArgValLeuThrTyrValHisAsnGluLeu	240	
DB	874 TGGATATTAACTGATACGTGTAACAGAAAGAAAGACCTCACTTATGTGCACAAATAGCA	933	
QY	241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu	260	
DB	934 GCCAACATGGAAGAAGATGCCAGATTCAGCTGAGTCCGCTTCTCCAAAGAA	993	
QY	261 PheAspArgTyrSerGlyTyrTyrTrpCysProLysAlaGluThrThrProSerGlyGly	280	
DB	994 TTTATATGATATTCCTGCTATTTGCTGTGTCCAAAGCTGAAACAACTCCAGATGCTGT	1053	
QY	281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleGluHisVal	300	
DB	1054 AAAATTTCTTGAATTTCTATATGAAAGAAATGATAAATCTGAGGTGGAAATTTATCATCT	1113	
QY	301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyTyr	320	
DB	1114 ACATGCCCTTATGTTGAAACAAAGAGGCGAGATTCAATTCCTTATCCAAACAGATGCA	1173	
QY	321 AlaAsnProLysValThrPheLysMetSerGluLeuMetIleAspAlaGluGlyArgIle	340	


```

Db      1174  GCAAATCCTAAAGTCACCTTTTAAGATGTCAGAAATATATGATTGATGCTGAAGAGATC 1233
Qy      341  ILeaspValIleaspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db      1234  ATAGATGTCATATGATAGAGAACTAATTCACCTTTTGAGATTCTATTGGAAGAGATTGAA 1293
Qy      361  TyrIleAlaArgAlaGlyTTrpThrProGluGlyLysTyrAlaTTrpSerIleLeuLeuAsp 380
Db      1294  TATATTGCGAGAGCTGAGTGAATGCTCTGAGGAAATATGCTTGAGTCCATCTCTACTAGT 1353
Qy      381  ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db      1354  CGCTCCAGACTCGCTCAGATAGTGTGATCTCACCGAATTATTATATCCAGATAGAA 1413
Qy      401  AspaPValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db      1414  GATGATGTTATGGAAGGACAGACTCATATGATGATGCTGATTCGTGTGACGCGACATA 1473
Qy      421  IleIleTyrGluGluThrThrAspIleTrrPileAsnIleIleAspIlePheHisValPhe 440
Db      1474  ATTATCTATGAAGAAACAACAGACATCTGATTAATATCCATGACATCTTTATGTTTTT 1533
Qy      441  ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db      1534  CCCCAAGTCACGAGAGAGAAATGAGTTATTTTGCTCTGATGCAAAACAGATTTC 1593
Qy      461  ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerGly 480
Db      1594  CGTCATTTATACAAAATATACATCTATTTAAAGAAAGCAATATATMAACGATCCGATGT 1653
Qy      481  GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db      1654  GGGCGCTGCTCCACAGTGAATTCAGAGTCTCTATCCAAAGAGAGATAGCAATTCACGT 1713
Qy      501  GlyIuTrrPgluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db      1714  GGTGAATGGGAAGTCTTGCGCGCATGATCTAATATCCAAAGTGAAGATCAGAAAG 1773
Qy      521  LeuValTyrPheGluGlyThrTyrAspSerProLeuGlnHisIleLeuTyrValValSer 540
Db      1774  CTGGTATATTTTGAAGGACCAAAAGATCCCTTTAGAGCATCACCTGTACCTACTAGT 1833
Qy      541  TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db      1834  TACGTAATTCCTCGAGAGGTGACAAAGCTGATGACCGTGGTATCTCATCTTGGCTGC 1893
Qy      561  IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db      1894  ATCAGTCAGCACTGTGACTTCTTTATAGTAAGTATAGTAACCAAGAAATCCACACTGT 1953
Qy      581  ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysValThrLysGluPhe 600
Db      1954  GTGTCCCTTTACAAAGCTATCAAGTCTTGAAAGTGAACCAACTGTGAAAAAAGAAATTT 2013
Qy      601  TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db      2014  TGGGCAACATTTTGATTCAGCAGAGTCTCTCTCTGACTATACCTCCCGAAATTTTC 2073
Qy      621  SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      2074  TCTTTTGAAGTACTACTGATTTTACATGTATGAGATGCTCTCAACACCTCATATCTA 2133
Qy      641  GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db      2134  CAGCTCGAAAGAAATATCTCTACTGCTGCTTCATATATGCGTGC----- 2180
Qy      661  ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      2180 ----- 2180
Qy      681  TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db      2180 ----- 2180

```

```

Qy      701  AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db      2181 -----GCTCAAAATGAAATTGACATCGATCGAGGAGGATCCAAATAT 2222
Qy      721  LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTTrpSerTyr 740
Db      2223  CTAGCTTCTCAATATGATTTATTTATTTGATTAATGATGATGAGGAGATCCAGGCTGCTAT 2282
Qy      741  GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db      2283  GAGAGATACCTCTCCCTGATGCAATTAATGACAGAGGTCAATATCTTCAGGGTGTCTATT 2342
Qy      761  AlaGlyAlaProValThrLeuTrrPilePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db      2343  GCTGGGCCCAAGCACACTCTGTGATCTTCTATGATATACAGATACACGGAACGTTATATG 2402
Qy      781  GlyIysProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db      2403  GGTACCTTGACCAAGATGAACAGGGCTATTAATCTTAGATCTGTGCGCATGCAACAGAA 2462
Qy      801  LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2463  AAGTTCCTCCCTGAAACCAATCGTTTACTGCTCTTACATGGTTTCTCGAGATGAGAAATGC 2522
Qy      823  HisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyrAsp 840
Db      2523  CATTTTGACATACACAGATATTTACTGATTTTATAGAGAGCGCTGGAACCATATGAT 2582
Qy      841  LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyr 860
Db      2583  TTACAGATCTATCTCTCAGAGACACACAGCATTAAGATCTCTGAATCGGAGAACATATAT 2642
Qy      861  GluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLys 880
Db      2643  GAATCGATCTTTTGACATCTTCAAGAAACCTTGATGATCAGATATTCGCTCTTAAA 2702
Qy      881  ValIle 882
Db      2703  GTGATA 2708

RESULT 8
US-09-976-674-8
; Sequence 8, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4523
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-976-674-8

Alignment Scores:
Pred. No.: 0
Score: 4092.50
Percent Similarity: 88.44%
Best Local Similarity: 88.44%
Query Match: 87.07%
DB: 10
Gaps: 1

US-10-070-464-1 (1-882) x US-09-976-674-8 (1-4523)

```

QY 1 MetAlaAlaMetGluThrGluGluGluValGlu1 ephGluThrAlaAspCys 20
 Db 214 ATGGCAGCGCATGGAAACAGACAGCTGGGTGGATATTTGAACTCGCGACTGT 273
 QY 21 GluGluAsn11eGluSerGluAspArgProLysLeuGluProPheTyrValGluArgTyr 40
 Db 274 GAGGAAATATTGAATCACAGGATCGGCTTAAATTGAGACCTTTTATGTGTAGCGGTAT 333
 QY 41 SerTPSerGluLeuLysLeuLeuAlaAspThrArgLysTyrHisGluTyrMetMet 60
 Db 334 TCTCGAGTCAGCTTAAAGCTGTCTTCCGATACAGAAATATCTATGCTTACATGAGT 393
 QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAspAspProAspGlyProHisSer 80
 Db 394 GCTAAGGACCACTATGATTTTCATGTTTGTGAAGAGAAATGATCCAGATGGACTCTTCA 453
 QY 81 AspArg11eTyrTyrLeuAlaMetSerGlyGluAsnArgLysAsnThrLeuPheTyrSer 100
 Db 454 GACAGAACTATTAACCTTCCATGTCTGGTGAGAAACAGAAATACACGTGTTTATCT 513
 QY 101 Glu11eProLysThr11eAspArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
 Db 514 GAAATTTCCAAACTATCAATAGAGCAGAGCTTAAATGCTCTTGGAAAGCTCTTTTGG 573
 QY 121 AspLeuPheGluAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
 Db 574 GATCTTTTTCAGGCAACACTGGACCTATGGAATGATATCTCGAAGAAAGAACTATTAA 633
 QY 141 GluArgLysArg11eGlyThrValGly11eAlaSerTyrAspTyrHisGluGlySerGly 160
 Db 634 GAAAGAAACGCAATGGAAACAGTCGGAAATTCCTTACGATATACCAAGAAAGTGA 693
 QY 161 ThrPheLeuPheGluAlaGlySerGly11eTyrHisValLysAspGlyGlyProGluGly 180
 Db 694 ACATTTCTGTTCAGCCGAGTGGATTTTATCACGTAAAGATGAGAGCCCAAGAA 753
 QY 181 PheThrGluGluProLeuArgProAsnLeuValGluThrSerCysProAsn11eArgMet 200
 Db 754 TTTACGAAACAACCTTTAAGGCCAATCTAGTGAAGAACTAGTTGCTCCAACTACGAG 813
 QY 201 AspProLysLeuCysProAlaAspProAspTrp11eAlaPhe11eHisSerAsnAsp11e 220
 Db 814 GATCCAAATATTGCGCTGCTGATCCAGACTGGATTCCTTTATACATGCAACGATATT 873
 QY 221 Trp11eSerAsn11eValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
 Db 874 TGGATATCTAACATGCTAAACAGAGAAAGAAAGAGACTCACTTATGTGACAAATGAG 933
 QY 241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGluGluGly 260
 Db 934 GCCAACATGGAAAGATGCCAGATCAGCTGGAAGTCCGTAACCTTTGTTCTCCAAAGAA 993
 QY 261 PheAspArgTyrSerGlyTyrTrpTyrCysProLysValGluThrThrProSerGlyGly 280
 Db 994 TTTGTAGATATTGCGCTATTGGTGTGTCCAAAGCTGAAACAACTCCCACTGGTGT 1053
 QY 281 Lys11eLeuArg11eLeuTyrGluGluAsnAspGluSerGluValGlu11eHisVal 300
 Db 1054 AAAATTTTAGAATTTATATAGAGAAATGATGAACTGAGGTGAATTAATTCATGTT 1113
 QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
 Db 1114 ACATCCCTATGTTGAAACAAAGAGGAGATTCATTCGTTATCTTAAACAGAGTACA 1173
 QY 321 AlaAsnProLysValThrPheLysMetSerGlu11eMet11eAspAlaGluGlyArg11e 340
 Db 1174 GCAATTCCTAAAGTACTTTTAAAGATGTCAAGAAATATGATGATGCTGAAGAGAGATC 1233
 QY 341 11eAspVal11eAspLysGluLeuLeu11eGlnProPheGlu11eLeuPheGluGlyValGlu 360
 Db 1234 ATAAGTCTATGATGAAAGAACTAATTCACCTTTTGAATTCATTTTGAAGAGATTGA 1293

QY 361 Tyr11eAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSer11eLeuLeuAsp 380
 Db 1294 TATATTGCCAGAGCTGATGAGACTCTGAGGGAATATGCTTGGTCCATCTCACTAGAT 1353
 QY 381 ArgSerGlnThrArgLeuGlu11eValLeu11eSerProGluLeuPhe11eProValGlu 400
 Db 1354 CCTCCCAACACTCGCTTACAGATAGTGTGATCTCACTGAATTTATTTTCCAGATGA 1413
 QY 401 AspAspValMetGluArg11eArgLeu11eGluSerValProAspSerValThrProLeu 420
 Db 1414 GATGATGTTATGAAAGCAGAGACTCATGATGCTGATCTGATTTCTGAGCCCACTA 1473
 QY 421 11e11eTyrGluGluThrThrAsp11eTrp11eAsn11eHisAsp11ePheHisValPhe 440
 Db 1474 ATTATCTAAGAAAGAAACAACAGACATCTGATTAATATCAATACACTTTTCACTGTTT 1533
 QY 441 ProGlnSerHisGluGluGlu11eGluPhe11ePheAlaSerGluCysLysThrGlyPhe 460
 Db 1534 CCCCAAGTCAGAAAGAGAAATTGAGTTTATTTTGGCTTGAATGCAAAACAGGTTTC 1593
 QY 461 ArgHisLeuTyrLys11eThrSer11eLeuLysGluSerLysTyrLysArgSerSerGly 480
 Db 1594 CGTCATTTATACAAATTAATCATTTTAAAGAAAGAAATATTAACGATCCAGTGT 1653
 QY 481 GlyLeuProAlaProSerAspPheLysCysPro11eLysGluGlu11eAla11eThrSer 500
 Db 1654 GGGCTCCCTCGCTCCAAGTATTTCAAGTCTCTATCAAGAGAGATAGCAATTACCACT 1713
 QY 501 GlyGluTrpGluValLeuGlyValArgHisGlySerAsn11eGluValAspGluValArgArg 520
 Db 1714 GGTGAATGGGAAGTTCTTGCCGCGCATGGATCAATATCAAGTTGATGAAGCAGAAAG 1773
 QY 521 LeuVal11eTyrPheGluGly11eThrLysAspSerProLeuGluHisGluLeuTyrVal11eSer 540
 Db 1774 CTGGTATATTTGAAAGCCCAAGACTCCCTTTAAGACATCACTGATACGATGACT 1833
 QY 541 TyrValAsnProGlyGlyValThrArgLeuThrAspArg11eTyrSerHisSerCysCys 560
 Db 1834 TAGTAAATCCCGAGAGGTGACAAAGCTGACATGACCGTGGCTACATCTTCTCTCC 1893
 QY 561 11eSerGlnHisCysAspPhePhe11eSerLysTyrSerAsnGlnLysAsnProHisCys 580
 Db 1894 ATCAGTGCAGCACTGACATCTTTTAAAGTATGATAGTAACAGAAAGATCCACACTGT 1953
 QY 581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
 Db 1954 GTGTCCCTTTACAACTATCAAGTCTGAAAGATGACCCCACTTGCACAAAGGAATT 2013
 QY 601 TrpAlaThr11eLeuAspSerAlaGlyProLeuProAspTyrThrProGlu11ePhe 620
 Db 2014 TGGGCCCACTTTTGGATTCAGCAGAGTCTCTTCTGACTATATCTCTCCAGAAATTTTC 2073
 QY 621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
 Db 2074 TCTTTTGAAGTACTAGTGAATTTTACATGTATGGATGCTCTCAAGCTCATGATCTTA 2133
 QY 641 GlnProGlyLysLysTyrProThrValLeuPhe11eTyrGlyGlyProGlnValGlnLeu 660
 Db 2134 CAGCTGGAAGAAATATCTCATCTGTGCTGATATAGTGT----- 2178
 QY 661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
 Db 2178 ----- 2178
 QY 681 TyrValValVal11eAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
 Db 2178 ----- 2178
 QY 701 AlaPheLysTyrLysMetGlyGln11eGlu11eAspAspGlnValGluGlyLeuGlnTyr 720
 Db 2178 ----- 2178
 QY 721 LeuAlaSerArgTyrAspPhe11eAspLeuAspArgValGly11eHisGlyTyrPheTyr 740

Db 2178 ----- 2178
QY 741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db 2179 ----- -CGGT- GCTATT 2189
QY 761 AlaGlyAlaProValThrLeuTrrIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db 2190 GCTGGGGCCCCAGTCACTGTGTGATCTTCTATGATACAGGATACACGGAACGTTATATG 2249
QY 781 GlyHisProAspGlnaengGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db 2250 GGTCAACCTGACACGAGATGAACAGGCTATTACTAGATCTGTGGCAGACGACGAA 2309
QY 801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db 2310 AAGTTCCTCCCTCAACCAATCGTTTACTCTCTTAATGATGTTCCCGGATGAGAAATGTC 2369
QY 821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db 2370 CATTTTGACATACCGATATATTAAGTATTTTACTAGAGGCTGGAAAGCCATATGAT 2429
QY 841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db 2430 TTACAGATCTATCTCTGAGAGAGACACAGCATAAAGATTCCTGATCGGAGAACATTAT 2489
QY 861 GluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db 2490 GAACCTGCATCTTTTGCACTACCTTCAGAAACCTTGATCAAGTATTCCTGCTCTAAAA 2549
QY 881 ValIle 882
Db 2550 GTGATA 2555

RESULT 9
US-09-976-674-14 ; Sequence 14, Application US/0976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: OI, Steve
APPLICANT: Aktinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 4309
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-14

Alignment Scores:
Pred. No.: 0 Length: 4309
Score: 3661.50 Matches: 708
Percent Similarity: 80.09% Conservative: 0
Best Local Similarity: 80.09% Mismatches: 1
Query Match: 77.90% Indels: 176
DB: 10 Gaps: 1

US-10-070-464-1 (1-882) x US-09-976-674-14 (1-4309)

QY 1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db 214 ATGGCAGCAGCATGGAAAGAAAGACAGCTGGGTCTTGAGATATTTGAAACTGGCGACTGT 273
QY 21 GluGluAsnIleGluSerGlnaAspArgProLysLeuGluProPheTyrValGluArgTyr 40

0
Db 274 GAGGAGAAATTTGAATCACAGGATCGGCTTAATTGGAGCTTTTATGTTAGCGGTAT 333
QY 41 SerTyrSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db 334 TCCTGAGATCAGCTTAAAAAGCTGCTGCCGATACCAAGAAAATATCATGGCTCATCATG 393
QY 61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db 394 GCTAAAGCACACATATATTCATGTTGTGMAAGAAATGATCAATGAGACCTCATTTCA 453
QY 81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db 454 GACAGATCTATTAATCTTGCCATGTCTGGTGAGAACAGAAATATACATGTTTATCTT 513
QY 101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrrLysProLeuLeu 120
Db 514 GAATTCCTCCAAACATATCAATAGACAGCAGCTTAATGCTCTTGGAAAGCTCTTTTG 573
QY 121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluLeuLeuArg 140
Db 574 GATCTTTTTCAGGCAACACTGACCTATGGAATGTATTTCTGAGAAAGAACTATTAGA 633
QY 141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db 634 GAAGAAAGAACGATTTGGAAACAGTCGGAATTCCTTTACATTAATACCAAGAAAGTGA 693
QY 161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db 694 ACATTTCTGTTTCAACCGGTAGTGAATTTATCAGTAAAGATGAGAGGCGACAAAGA 753
QY 181 PheThrGlnGluProLeuArgProAsnLeuValGluThrSerGlyProAsnIleArgMet 200
Db 754 TTACGCCAACACCTTTAAAGCCCAATCTAGTGAATCTAGTTGCTCCCAACATACGAGT 813
QY 201 AspProLysLeuGlyAspProAlaAspProAspTrrIleAlaPheIleHisSerAsnAspIle 220
Db 818 GATCCAAATTAATGCGCTGCTGTATCCAGACTGATGATGCTTTTATACATAGCAACGATAT 873
QY 221 TrrIleSerAsnIleValThrArgGluGluArgGluLeuThrTyrValHisAsnGluLeu 240
Db 874 TGGATATCTTAACATCGTAACAGAGAAAGAAAGAGACTCATATGTCACATAGACTA 933
QY 241 AlaAsnMetGluGluAsnAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db 934 GCCAACATGGAAAGAAATGCGCAGATCAGCTGAGCTGCTACCTTGTCTCCAAAGAA 993
QY 261 PheAspArgTyrSerGlyTyrTrrPrrCysProLysAlaGluThrThrProSerGlyGly 280
Db 994 TTTGATAGATATTTCTGCTATGCTGTGTCCAAAGCTGAAACACTCCCACTGGTGTGT 1053
QY 281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db 1054 AAAATCTTAAGATTTCTATATGAAAGAAATGATGATCTGAGGTGAAATTTATTCATG 1113
QY 301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrArgTyr 320
Db 1114 ACATCCCTATGTTGAAACCAAGAGGCGAGATCTTCCGTTATCTTAAACAGGTACA 1173
QY 321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgGly 340
Db 1174 GCAAAATCCTAAGTCTCTTTAAAGATGTCAGAAATATATGATGATCTGAAAGAAATATC 1232
QY 341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db 1232 ----- 1232
QY 361 TyrIleAlaArgAlaGlyTrrPrrProGluGlyLysTyrAlaTrrSerIleLeuLeuAsp 380
Db 1232 ----- 1232
QY 381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400

```

Db      1232 ----- 1232
Qy      401 AspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThrProLeu 420
Db      1232 ----- 1232
Qy      421 IleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisValPhe 440
Db      1232 ----- 1232
Qy      441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIleThrGlyPhe 460
Db      1232 ----- 1232
Qy      461 ArgHisLeuTyrIleThrSerIleLeuLysGluSerLysTyrLysAspSerSerGly 480
Db      1232 ----- 1232
Qy      481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db      1232 ----- 1232
Qy      501 GlyGluTyrGluValLeuGluYargHisGlySerAsnIleGlnValAspGluValArgArg 520
Db      1233 ----- 1253
Qy      521 LeuValTyrPheGluGluTyrHisLysAspSerProLeuGluHisIleLeuTyrValValSer 540
Db      1254 CTGGATATATTTGAAAGCACCMAAGACTCCCTTTAGAGCATGACCTGATGCAAGTACAGTACAGT 1253
Qy      541 TyrValAsnProGluValValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db      1314 TACGTAAATCTCGAGAGAGGTGACAGAGCTGACGACCGTGGCTACCTGACATTTCTTCTCTG 1373
Qy      561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db      1374 ATCAGTCACACACTGACTCTTTATAGTAAGTAATAGTAACAGAAAGATCCACACTGT 1433
Qy      581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db      1434 GTGTCCTTTACAGAGTATCAAGTCTCTGAAGATGACCCCACTTGCAAAACAAAGAAATTT 1493
Qy      601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePhe 620
Db      1494 TCGGGCACCATTTTGGATTGACAGAGGCTCTTCTGACATATCTCTCCAGAAATTTTC 1553
Qy      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      1554 TCTTTGAAAGTACTACGTGATTACATTGTATGGAGTGTCTTACCAAGCCTCATGATCTA 1613
Qy      641 GlnProGluLysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnL 660
Db      1614 CAGCCTGGAAAGAAATATCTCTACTGTGCTGTCATATATGAGTGCTCTCCACAGTGCAGT 1673
Qy      660 euValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuG 680
Db      1674 TCGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTCAGCTCTCTAG 1733
Qy      680 LysTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluG 700
Db      1734 GTTATGTGGTGTAGATAGACAAACAGGGGATCCCTGTACCGAGGGCTTTAAATTTGAAG 1793
Qy      700 LysAlaPheLysTyrLysMetGlyGlnIleGlnIleAspAspGlnValGluGlyLeuGlnT 720
Db      1794 GCGGCTTTAAATATTAATAGGGTCAATATGAATTAACATCAAGGAGGAAAGCATCTCAAT 1853
Qy      720 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyLysIleGlyTyrSerT 740
Db      1854 ATCTAGACTTCTCGATATGATTTCACTGACTTATGATGTGTGGGCACTCCAGGCTGTCTCT 1913
Qy      740 TyrGlyGlyTyrLysSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760
Db      1914 ATGGAGGATACCTCTCCCTGATGGCATTAATATCAGAGGTCAATATCTTCAGAGTTGGCTA 1973

```

```

Qy      760 LeAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrM 780
Db      1974 TTCTGGGGCCCAAGTACACTCTGTGATCTTTCTATGATACAGATACAGAACTTATATA 2033
Qy      780 eTGHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaG 800
Db      2034 TGGGTCACTCCAGTCAAGAAATGAAACAGGGCATATTAGGATCTGTGGCCATCAAGCAG 2093
Qy      800 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnV 820
Db      2094 AAAAGTTCCTCTGAAACCAATCGTTACTGCTTACATGGTTCTCTGAGAGAGATG 2153
Qy      820 AlaHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrA 840
Db      2154 TCCATTTTGACATACACAGATATTAATCTAGATTTTATGAGAGGGCTGGAAAGCCATATG 2213
Qy      840 sPLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisT 860
Db      2214 ATTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTCCTGAAATCGGAGAACATT 2273
Qy      860 YrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeu 880
Db      2274 ATGAACGTGATCTTTTGCACTACCTTCAAGAAACCTTGGATCAGTATTGCTGCTTAA 2333
Qy      880 YrValIle 882
Db      2334 AAGTGATA 2341

RESULT 10
US-09-976-674-4
; Sequence 4, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2617
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-4
Alignment Scores:
Pred. No.: 0 Length: 2617
Score: 2870.00 Matches: 517
Percent Similarity: 77.50% Conservative: 134
Bec Local Similarity: 61.55% Mismatches: 187
Query Match: 61.06% Indels: 2
DB: 10 Gaps: 2
US-10-070-464-1 (1-882) x US-09-976-674-4 (1-2617)
Qy      35 PheTyrValGluArgTyrSerTyrPheGlnLeuLysLysLeuLeuAlaAspThrArgLys 54
Db      80 TTCAGAGTGCAGAGAGACTCTGTGGACGCGGCTCCGAGCATCATCCACGCGACCCGCAAG 139
Qy      55 TyrHisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74
Db      140 TACTCGGCGCTCATCTTCAACAAAGGCGCCCGACGACTTTCAGATTGTGCAAGAGAGAGAT 199
Qy      75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGlu 94
Db      200 GAGTCTGGGCGCCCACTCCACGCGCTTACTACTGTGGAAATGCATATATGCGAGCGGAGAG 259

```


Db 2420 CACGGCTTCTGAGCAAAAAACGTGCACTTTTCCACAAACACTTCTGCTCCCAACTG 2479
QY 833 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852
Db 2480 ATCCGAGCAGGAGAAACCTTACACGACTCTACAGTCCCAACGAGAGACACACTATTCCG 2539
QY 853 ValProGlnSerGlyGlnIleTyrGlnIleLeuHisIleTyrLeuGlnGlnIleuLeu 872
Db 2540 TGCCCCGAGTCCGAGCAAGCACTATGAAATGACGTTGCTGACACTTCTTACAGGAATACCTC 2599
RESULT 11
US-09-976-674-28
Sequence 28, Application US/09976674
Patent No. US2002015843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinaanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28
LENGTH: 4219
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-28
Alignment Scores:
Pred. No.: 0 Length: 4219
Score: 2870.00 Matches: 517
Percent Similarity: 77.50% Conservative: 134
Best Local Similarity: 61.55% Mismatches: 187
Query Match: 61.06% Gaps: 2
DB: 10
US-10-070-464-1 (1-882) x US-09-976-674-28 (1-4219)
QY 35 PheTyrValGluArgTyrSerTyrPsrSerGlnLeuLysLeuLeuAlaAspThrArgLys 54
Db 436 TTCGAGTGCAGAGCACTCGTGGACGGGCTCGGAGCATTCACACGCGACGCCCAAG 495
QY 55 TyrHisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValIlyAspAsp 74
Db 496 TACTCGGGCTCATTTGTCAACAAGGCGGCCCAAGCTTCACAGTTTGTGCAAGAGCGAT 555
QY 75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyLysAspArgLys 94
Db 556 GAGTCGGGCCCCCACTCCACCGCTCTTACTCTGGAATGCCATATGGCCGAGAG 615
QY 95 AsnThrLeuPheTyrSerGlnIleProLysThrIleAsnArgAlaIleValLeuMetLeu 114
Db 616 AACTCCCTCTCTACTCTAGATTCCCAAGAAAGTCCGGAAGAAGGCTCTGCTGCTCCG 675
QY 115 SerTyrLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
Db 676 TCTCTGAACAGATGCTGATCATTTCCAGGCCACCCCAACATGGGCTTACTCTCG 735
QY 135 GlnGlnGlnLeuLeuArgLysArgLysArgIleGlyThrValGlyIleAlaSerTyrAsp 154
Db 736 GAGGAGACCTGCTGAGGAGCGGAAACCTCGGGGCTTCGCGCATCACCTCTAGAC 795
QY 155 TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174
Db 796 TTCACAGGAGAGTGGCTCTTCTCTCCAGGCCACGACACGCTTCCACATGCCCG 855
QY 175 AspGlyGlyProGlnIlePheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194

Db 856 GACGGCGCAAGAACGGCTTCACTGCTGCTCCCTATGAAACCGCTGGAATATCAAGACCAG 915
QY 195 CysProAsnIleArgMetAspProLysLeuCysProAlaAspProAspTrpIleAlaPhe 214
Db 916 TGCTCAGGCGCCCGGATGAGACCCCAAAATCTGCCCTGCGACCTGCTCTTCTCCTCC 975
QY 215 IleHisSerAspAspIleTyrPheSerAsnIleValThrArgGlnGluArgGlyLeuThr 234
Db 976 ATCAATTAACAGCACTGTGGGTGGCCAAATCATGACACAGGACGAGCGGCGGTGACC 1035
QY 235 TyrValHisAsnGlnLeuAlaAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThr 254
Db 1036 TCTGCGCAACAGCTTATCCATATGCTCCGATGAGACCCCAAGCTCGGGGTGGCCACC 1095
QY 255 PheValLeuGlnGlnLeuPheAspArgTyrSerGlyTyrTrpCysProLysAlaGlu 274
Db 1096 TTGCTCATACAGGAAGATTCGACCCGCTTCACTGCGGTACTGCTGCTCCCAAGCTCC 1155
QY 275 ThrThrProSerGlyLys---LysIleLeuArgIleLeuTyrGlnGlnIleuAspGlnSer 293
Db 1156 TGGGAAGTTTCAGAGGCTCCAAAGACGCTGCGAATCTGTATGAGAAAGTCGATGAGTCC 1215
QY 296 GlnValGlnIleIleHisValThrSerProMetLeuGlnThrArgArgAlaAspSerPhe 313
Db 1216 GAGGTGAGGTATTCATCAGTCCCTCTGCGCTGAGAAAGAAAGACGACTCTGAT 1275
QY 314 ArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGlnIleMet 333
Db 1276 CGGTACCCAGGACAGGACGAAAGATCCCAAGATTGCTTGAATCGGCTGAGTTCAC 1335
QY 334 IleAspAlaGlnGlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGln 353
Db 1336 ACTGACAGCGAGGCAAGATCTGCTGACCCAGCAAGAGAGCTGTGCGAGCCCTTCAAC 1395
QY 354 IleLeuPheGlnGlyValGlnTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyr 373
Db 1396 TCGCTTCTCCAAAGTGAATGATTCGCGAGGCGGGGTGACCCGGATGCGAAATAC 1455
QY 374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393
Db 1456 GCGTGGCCATGTTCTGAGACCGGCCAGAGTGGCTCAGCTGCTCTCCGCCCG 1515
QY 394 GlnLeuPheIleProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerVal 413
Db 1516 GCCCTGTATCCCGGACACAGAAATGAGAGACGGCTAGCCTCTGCCAGACGCTC 1575
QY 414 ProAspSerValThrProLeuIleIleTyrGlnGlnThrThrAspIleTrpIleAsnIle 433
Db 1576 CCCAGGAATGTCAGCCGATATGCTGATCGATGAGAGAGTACCAACCTTGATCATGTT 1635
QY 434 HisAspIlePheHisValPheProGlnSerHis--GlnGlnGlnIleGlnPheIlePhe 452
Db 1636 CATGACATCTCTATCTCTCCCAATCAGAGGAGAGAGACGAGCTGCTTCTCCCG 1695
QY 453 AlaSerGlnCysLysThrGlyPheArgHisIleLeuTyrLysIleThrSerIleLeuLysGln 472
Db 1696 GCCAATGAATGCAAGACCGGCTTTCGCAATTGTATCAAAATCCGCCGTTTAAATCC 1755
QY 473 SerLysTyrLysArgSerSerGlyLysProAlaProSerAspPheLysCysProIle 492
Db 1756 CAGGCTTACGATTTGAGTACGCCCTTACGCCCGGGGAAGATTAATTAAGTCCCAAT 1815
QY 493 LysGlnGlnIleAlaIleThrSerGlyGlnTyrGlnValLeuGlyArgHisGlySerAsn 512
Db 1816 AAGGAAGAGATTTGCTCTGACAGCGGTGATGAGAGTTTGGCGAGCGACGCTCCAG 1875
QY 513 IleGlnValAspGlnValArgArgLeuValTyrPheGlnGlyThrLysAspSerProLeu 532
Db 1876 ATCTGGGTCAATGAAGAGACCAAGCTGTGTACTTCCAGGACCAAGACACGCGCTG 1935
QY 533 GlnHisIleLeuTyrValLysTyrValAsnProGlnGlnValThrArgLeuThrAsp 552

Db 1936 GAGCACCACCTTACGTGTACGTATAGAGCGGCGGAGATCGTACCGCTCACCACG 1995
Qy 553 ARGGLTYRserHisserCysValleSerGlnHisCysAspPhePheIleSerLysTYR 572
Db 1996 CCCGGCTCTCCATAGACTCTCCATAGCCAGAACTTCACATGTTCTGTCAGCCTAC 2055
Qy 573 SerAsnGlnLysAsnProHisCysValleSerLysTYRserLysSerProGluAspAsp 592
Db 2056 AGCAGCGTGAGCAGCGCGCGCTGCTGTCAGCTCAGAGCTGAGCGCGCCCGAGCAGAC 2115
Qy 593 ProThrCysLysThrLysGluPheTPAAlaThrIleLeuAspSerAlaGlyProLeuPro 612
Db 2116 CCCCTGCAGCAGAGCCCGCTCTGCGCTGAGTATGAGCAGCAGCCTGCCCCCG 2175
Qy 613 AspThrThrProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTYRGLY 632
Db 2176 GATTATGTTCTCCAGATCTTCATTCCACACCGCGCTGGATGTGCGGCTCTAGGCG 2235
Qy 633 MetLeuTYRAspProHisAspLeuGlnProGlyLysLysTYRProThrValleuPheIle 652
Db 2236 ATGATCTACAAAGCCCGCCTTGCGACCGAGGAGAGAGACCCCGCTCTTTGTA 2295
Qy 653 TYRGLYGLYProGlnValleuLeuValleuAsnAspPheLysGlyValLysTYRPhenArg 672
Db 2296 TATGAGGCGCCCGAGGTGAGCTGTGTAATACCTCTTCAAAAGCATCAAGTACTGCGG 2355
Qy 673 LeuAsnThrLeuAlaSerLeuGlyTYRValValleuValleuAspAsnArgLysSerCys 692
Db 2356 CTCACACACTGGCTCTCTCGGCTACGCGCTGTGTGATGACGCGCAGGCGCTCTGT 2415
Qy 693 HisArgGlyLeuLysPheGluGlyAlaPheLysTYRLeuMetGlyGlnIleGluIleAsp 712
Db 2416 CACGAGGCGCTCGGTTCCAAAGGCGCCCTGAAAAACCAATGGGCGCAGTGGAGATCGAG 2475
Qy 713 AspGlnValleuGlnLysLeuGlnTYRLeuAlaSerArgTYRAspPheIleAspLeuAspArg 732
Db 2476 GACCAAGTGGAGGCGCTGCGATGCGGCGGAGAGATGCTTCATCGACCTGAGCGCA 2535
Qy 733 ValGlyIleHisGlyTYRTPSerTYRGLYGLYTYRLeuSerLeuMetAlaLeuMetGlnArg 752
Db 2536 GTTGCCATCGATGGCTGTGCTCTACGGGCGCTCTCTCGCTCATGGGGCTTATTCACAG 2595
Qy 753 SerAspIlePheArgValleuAlaIleAlaGlyAlaProValThrLeuTPR1lePheTYRAsp 772
Db 2596 CCCCAAGTGTTCAAAGTGGCCATCGCGGCTGCGCGCTCAGCTGAGTGGCTTACGAC 2655
Qy 773 ThrGlyTYRThrLysArgTYRMetGlyHisProAspGlnAsnGlnGlnGlyTYRLeu 792
Db 2656 ACAGGGTACACTGAGCGCTACATGAGACGTCTCGAGAAACACAGCAGCGCTATGAGGCG 2715
Qy 793 GlySerValleuMetGlnAlaGlnLysPheProSerGluProAsnArgLeuLeuLeu 812
Db 2716 GGTTCCTGTGCGCTGACCGTGGAGAGCTGCCCAATGAGCCCAACCGCTTATTCCTC 2775
Qy 813 HisGlyPheLeuAspGlnValleuHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
Db 2776 CACGGCTTCTCGAGCAAAACGTGCACTTTTCCACAACTTCCTCGCTCCCAACTG 2835
Qy 833 ValArgAlaGlyLysProTYRAspLeuGlnIleTYRProGlnLysArgHisSerIleArg 852
Db 2836 ATCCGAGCAGGAGAACTTACAGCTCAGATCTACCCCAACGAGAGACCACTATTTCGC 2895
Qy 853 ValProGluSerGlyGlnHisTYRGLYLeuHisLeuLeuHisTYRLeuGlnGlnValleu 872
Db 2896 TGCCTCGAGTCCGGCGAGGACCTATGAAAGTCAAGTGTGCACTTTCTACAGAAATACCTC 2955

RESULT 12
US-09-976-674-24

; Sequence 24, Application US/0997674

; Patent No. US20020115843A1

; GENERAL INFORMATION:

; APPLICANT: OJ, Steve

; APPLICANT: Akinsanya, Karen

; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTI
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 4302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-24
Alignment Scores:
Pred. No.: 0 Length: 4302
Score: 2870.00 Matches: 517
Percent Similarity: 77.50% Conservative: 134
Best Local Similarity: 61.55% Mismatches: 187
Query Match: 61.06% Indels: 2
DB: 10 Gaps: 2
US-10-070-464-1 (1-882) x US-09-976-674-24 (1-4302)

Qy 35 PheTYRValleuArgTYRserTPSerGlnLeuLysLeuLeuAlaAspThrArgLys 54
Db 436 TTCAGAGCAGAGAGCACTGTGGAGCGGCTCCGAGCATCTCCACGCGACCGCAG 495
Qy 55 TYRHisGlyTYRMetMetAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74
Db 496 TACTCGGCTCATTTCTTCAAGAGCGCGCCACGACTTCAGTGTGTGCGAAGAGCGAT 555
Qy 75 ProAspGlyProHisSerAspArgIleTYRLeuAlaMetSerGlyGluAsnArgGlu 94
Db 556 GAGTCTGGGCGCCCACTCCACCGCTCTACTACCTGGAAATGCATATGGCAGCCGAG 615
Qy 95 AsnThrLeuPheTYRserGlnIleProLysThrIleAsnArgAlaValleuMetLeu 114
Db 616 AACTCCCTCTCTACTGTGATTCCTCAAGAGTCCGAGAAAGCTCGTGTCTCTG 675
Qy 115 SerTPR1ProLeuLeuAspLeuPheGlnAlaThrLeuAspTYRGLYMetTYRserArg 134
Db 676 TCTTGAGACAGATGTGTGATATTTCCAGGCCACGCCACCATGGGTCTACTCGG 735
Qy 135 GlnGlnGlnLeuLeuArgLysArgIleGlyTYRValGlyIleAlaSerTYRAsp 154
Db 736 GAGGAGAGACTGTGAGGAGCGGAAACCTGTGGGCTCTTGGCATCACCTCTAGAC 795
Qy 155 TYRHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTYRHisValLys 174
Db 796 TTCACAGCGAGAGTGGCTCTTCTCTTCCAGGCGCAGAACGCTCTTCCACTCGCGC 855
Qy 175 AspGlyGlyProGlnIlePheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194
Db 856 GAGGGGGAGAGAACGCTTCAATGTGTGCTCCCTATATAAACCGTGGAAATCAAGACCG 915
Qy 195 CysProAsnIleArgMetAspProLysLeuCysProAlaAspProAspTPR1leAlaPhe 214
Db 916 TGTCTAGGCGCCCGGATGAGACCCCAAAATCTGCGCTGCGGACCTCTTCTCTCCTC 975
Qy 215 IleHisSerAsnAspIleTPR1leSerAsnIleValThrArgGlnLysArgLysLeuThr 234
Db 976 ATCAATTAACAGAGCACTGTGGTGGCGCAATCGAGACAGCGGAGAGCGCGCTGACC 1035
Qy 235 TYRValHisAsnGlnLeuAlaAsnMetGlnLysAspAlaArgSerAlaGlyValAlaThr 254
Db 1036 TTCTGCCACCAAGGTTTATCCATGTCCCGATGAGCCCAAGTCTCGGGTGTGGCCACC 1095
Qy 255 PheValleuGlnGlnLysPheAspArgTYRserGlyTYRTPR1CysProLysAlaGlu 274
Db 1096 TTCTGCATACAGAGAGATTCAGCCGCTTCACTGGGTACTGTGTGCGCCACAGCTTCC 1155

QY 275 ThrThrProSerGlyGly---LysIleLeuArgIleLeuTyrgluGluLeuAspGluSer 293
 DB 1156 TGGGAAGTTCAGAGAGCCCTCAAGACGCTCCGAATCTGTGATGAGAAAGCATGAGTCC 1215
 QY 294 GluValGluIleIleHisValThrSerProMetLeuGluThrArgAlaAspSerPhe 313
 DB 1216 GAGGTGAGGTGATCATCAGCTCCCTCTCTGCGCTGAGAAAGAAAGAACGAGCATGTAT 1275
 QY 314 ArgTygProlyserThrLysAlaAsnProlyValThrPheLysMetSerGluIleMet 333
 DB 1276 CGGTACCCAGGACAGGACAGCAAGAAATCCCAAGATTCCTTGAACATGGCTGTGATCCAG 1335
 QY 334 IleAspAlaGluGlyArgIleIleAspValIleAspGluLeuLeuIleGlnProPheGlu 353
 DB 1336 ACTGACAGCGAGGCAAGATCGTCTGACCCAGAGAAAGAGTGTGTGAGCCCTTCAGC 1395
 QY 354 IleLeuPheGluGlyValGluTygIleAlaArgAlaGlyTrpThrProGluGlySerTyg 373
 DB 1396 TCGCTGTTCCTCCAGAGGTGAGTACATCGCCAGGGCGGGTGGACCCGGATGGCAATAC 1455
 QY 374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleuIleSerPro 393
 DB 1456 GCTTGAGCATGTTCTTGAGACGGGCCAGACAGTGGCTCAGCTCGTCTCTCCGCCCG 1515
 QY 394 GluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413
 DB 1516 GCCCTGTTCATCCCGACAGCAAGATGAGAGACGGGTAGCCCTCCAGAGCTGTTC 1575
 QY 414 ProAspSerValThrProLeuIleIleTyrgluGluThrThrAspIleTrpIleAsnIle 433
 DB 1576 CCAGAGAAATGTCAGCCGTATGTGTGTACGAGAGAGTACCAACGTCGTGATCATGT 1635
 QY 434 HisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePhe 452
 DB 1636 CATTGACATTTCTATCCCTTCCTCCCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695
 QY 453 AlaSerGluCysLeuThrGlyPheArgHisLeuTygIleIleThrSerIleLeuLysGlu 472
 DB 1696 GCCAATGATGACAGACGGCTTCTGCTCCATTGTACAAAGTCAACGCCCTTTAAATCC 1755
 QY 473 SerLysTygLeuAspArgSerSerGlyLeuProAlaProSerAspPheLysCysProIle 492
 DB 1756 CAGGCTCAGATTGGAGTGAGGCCCTTCAACCCCGGGGAGAGATGAAATTTAAGTCCCAT 1815
 QY 493 LysGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAsn 512
 DB 1816 AAGGAAGATTTGCTCTGACAGCGGTGATGAGAGTTTGGCGGAGGACGAGCTTCCAA 1875
 QY 513 IleGlnValAspGluValArgArgLeuValTygPheGluGlyThrLysAspSerProLeu 532
 DB 1876 ATCTGGGTCAATGAGAGAGACCAAGCGTGTACTTCCAGGGACCAAGACACGCCGCTG 1935
 QY 533 GluHisHisLeuTygValValSerTygValAsnProGlyGluValThrArgLeuThrAsp 552
 DB 1936 GAGGACCAACCTCTACGTGTGAGCTATGAGCGCGCGGAGATGTGAGCGCTCAGCAG 1995
 QY 553 ArgGlyTygSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerTyg 572
 DB 1996 CCCGCTTCTCCATAGCTGCTGCTCCATGAGCCCAAACTTGACATGTTGCTCAGCACTAC 2055
 QY 573 SerAsnGlnLysAsnProHisCysValSerLeuTygLeuSerSerProGluAspAsp 592
 DB 2056 AGCAGCGTAGAGACCGCGCGCTGCGTACCGTCAAGCTGAGGCGGCCGACAGCAGAC 2115
 QY 593 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 612
 DB 2116 CCCCTGCAACAGCACCGCGCTTCTGGGTACATGATGAGAGCAGCAGCTGCCCGCG 2175
 QY 613 AspTygThrProProGluIlePheSerPheGluSerThrGlyPheThrThrLeuTyg 632
 DB 2176 GATTATGTTCTCCAGAGATCTTCATTTCACAGCGCGCTGAGATGTGGCGCTTCAAGCG 2235

QY 633 MetLeuTygIleAspProHisAspLeuGlnProGlyLysLeuTygProThrValLeuPheIle 652
 DB 2236 ATGATCTCAAGGCCCAACGCCCTTGGACGAGGAGAAAGACCCCAACGCTCTTGTGA 2295
 QY 653 TygIleGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTygPheArg 672
 DB 2296 TATGAGGCCCCCAGAGTGGAGGCTGTGAATTAACCTTCAAGAGCATCAAGTACTTGGCG 2355
 QY 673 LeuAsnThrLeuAlaSerLeuGlyTygValValValIleAspAsnArgGlySerCys 692
 DB 2356 CTCACACACTGGCTCCCTCGGCTTACGCGGTGTGTGATTCACGCGAGGGGCTCTGT 2415
 QY 693 HisArgGlyLeuLeuPheGluGluValAlaPheLysTygLeuMetGlyGlnIleGluIleAsp 712
 DB 2416 CACGAGGCTTCCGTTCCAGAGGAGCCCTGAAACCAACAAAGGAGGAGGAGATTCAG 2475
 QY 713 AspGlnValGluGlyLeuGlnIleValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTygAsp 732
 DB 2476 GACAGAGTGAAGGCTTGACGATTGCGGCGGAGAAATATGCGCTTATCCAGCTGAGCCGA 2535
 QY 733 ValGlyIleHisGlyTrpSerTygIleTygIleLeuSerLeuMetAlaLeuMetGlnArg 752
 DB 2536 GTTGCATTCATGAGCTGCTGCTTACAGGGGCTTCTCTGCTCATGAGGGCTAATCCACAG 2595
 QY 753 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTygAsp 772
 DB 2596 CCCAGGTTCATAGGTGACCATTCGCGGTGCGGCTCCGCTGATGAGCTTACGAC 2655
 QY 773 ThrGlyTygThrGluArgTygMetGlyHisProAspGlnAsnGluGlnIleTygTygLeu 792
 DB 2656 ACAGGGTACACTGAGGCTTACATGACGTCCTGAGAACACACAGACGGCTATGAGGCG 2715
 QY 793 GlySerValAlaMetGlnAlaGluLysPheProSerGlnProAsnArgLeuLeuLeu 812
 DB 2716 GGTTCGTCGCTCGCTGACCTGAGAGAGCTGACCAATGAGCCCAACCGCTTGTATCTTC 2775
 QY 813 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
 DB 2776 CACGGCTTCTGAGAGAAACGACGCTTTTCCACAAACTTCTCGTCTCCAACTG 2835
 QY 833 ValArgAlaGlyLysProTygAspLeuGlnIleTygProGlnGluArgHisSerIleArg 852
 DB 2836 ATCCGAGCAGGAAACCTTACCAAGCTCCAGATCTACCCCAAGAGAGACAGCATATTCC 2895
 QY 853 ValProGluSerGlyGluHisTygIleLeuHisLeuLeuHisTygLeuGlnGluAsnLeu 872
 DB 2896 TCCCGCGAGTCGGCGAGGACATATGAAATGACGTTGCTGCTCTACAGAGATACCTC 2955

RESULT 13

US-09-976-674-36
 / Sequence 36, Application US/09976674
 / Patent No. US20020115843A1
 / GENERAL INFORMATION:
 / APPLICANT: Qi, Steve
 / APPLICANT: Akinsanya, Karen
 / APPLICANT: Riviere, Pierre
 / APPLICANT: Junien, Jean-Louis
 / TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
 / FILE REFERENCE: 70669
 / CURRENT APPLICATION NUMBER: US/09/976,674
 / PRIOR FILING DATE: 2001-10-12
 / PRIOR APPLICATION NUMBER: US 60/240,117
 / NUMBER OF SEQ ID NOS: 61
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 36
 / LENGTH: 4180
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 US-09-976-674-36

Alignment Scores: 0 Length: 4180
 Pred. No.: 0

Score: 2820.50 Matches: 510
Percent Similarity: 76.43% Conservative: 132
Best Local Similarity: 60.71% Mismatches: 183
Query Match: 60.01% Indels: 15
DB: 10 Gaps: 3

US-10-070-464-1 (1-882) x US-09-976-674-36 (1-4180)

Qy PheTyrValGluArgTyrSerGlnLeuLysLeuAlaAspThrArgLys 54
Db TTTCCAGTGCAGAGCACTCGTGGACGGCTCCGAGCATATCACCGCAGCGCAAG 495
Qy TyrHisGlyTyrMetAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74
Db TACTCGGGCCCTCATTTGTCACAGCGCGCCACGACTTCGATTGGTCAGAAACAGCAT 555
Qy ProAspGlyProHisAspArgLysLeuTyrTyrLeuAlaMetSerGlyLysAsnArgLys 94
Db GAGTCTGGGGCCCACTCCACCGCTCTACTACTGGAATGCCATATGGCAGCCGAGAG 615
Qy AsnThrLeuPheTyrSerGlnLeuLeuProLysThrIleAsnArgAlaAlaValLeuMetLeu 114
Db AACCTCCCTCTCTACTCTGAGATTCACAGAAAGTCCGAAAGAGCTCTGCTCTCTG 675
Qy SerTyrLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
Db TCTCGAAGACGATGCTGATCATTTCCAGGCCACGCCCATGGGGTCTACTCTCGG 735
Qy GlnGlnGlnLeuLeuArgLysArgLysArgLysArgLysArgLysArgLysArgLys 154
Db GAGAGAGAGCTGCTGAGGAGCGGAAACGCTGGGGTCTTGGCATCACCTCTACGAC 795
Qy TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174
Db TTCCACAGCCAGATGGCTCTCTCTCTCCAGGCCACGAAACACCTCTTCCATCGCCG 855
Qy AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194
Db GACGCGCGCAAGACGGCTTCATGCTGCTCTATGAAACCGCTGAAATCAAGACCCAG 915
Qy CysProAsnIleArgMetAspProLysLeuCysProLysProLysProLysProLys 214
Db TGCTCAGGGCCCCGGAATGAGCCCAAAATCTGCCCCCTGCACTCTCTCTCTCTCT 975
Qy IleHisSerAsnAspIleTyrPheLysSerAsnIleValThrArgGlnGluArgLysLeuThr 234
Db ATCAATAACAGCGACCTGTGGTGGCCAACTCAGACAGCGCAGCGCGCTACCC 1035
Qy TyrValHisAsnGlnLeuAlaAsnMetGlnGluAspAlaArgSerAlaGlyValAlaThr 254
Db TTCTGCCACCAAGGATTATCCAAATGCTCTGGATGACCCCAAGTCTGGGGGTGGCCACC 1095
Qy PheValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 274
Db TTTCTCTATACAGAAAGTCTGACCGCTTCACTGGGATCTGGTGGTGGTGGTGGTGG 1155
Qy ThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlnGluAsnAspGlyLys 293
Db TGGGAAGCTTCAGAGGGCTCAAGACCGCTGGATCTCTATATGAGAAAGTCGATGCC 1215
Qy GluValGlnIleIleHisValThrSerProMetLeuGluThrArgAlaAspSerPhe 313
Db GAGGTGAGGTCAATTCAGTCCCTCTCTGCTGCTCAAGAAAGAAAGAAAGAAAGAAAG 1275
Qy ArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGlyIleMet 333
Db CGGTAACCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1335
Qy IleAspAlaGlnGlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGln 353
Db ACTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1395
Qy IleLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 373

Db TCGCTGTTCGGAAGTGCAGTACATGCCAGCGCCGGTGGACCCGGGAATGGCAATAC 1455
Qy AlaTyrSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleLysPro 393
Db GCGTGGGCGCATGTTCTCGACCGCGCCACAGTGGCTCCAGCTCTCTCTCTCTCTCT 1515
Qy GluLeuPheIleProValGluAspAspValMetGlnArgGlnArgLeuIleGluSerVal 413
Db GCCTGTTCATCCGACACAGAAATGAGAGAGCGGCTAGGCTCTGCCAGAGCTGTTC 1575
Qy ProAspSerValThrProLeuIleIleTyrGlnGluThrThrAspIleTyrPheAsnIle 433
Db CCGAGAAATCCAGCGCGTATGTGTATAGAGAGGTACCAACGCTGGATCAATGTT 1635
Qy HisAspIlePheHisValPheProGlnSerHisGlnGlnGlnIleGluPheIlePhe 452
Db CATGACATCTTATATCCCTTCCCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695
Qy AlaSerGluCysLysThrGlyPheArgHisLeuTyrIleValThrSerIleLeuLysGlu 472
Db GCCAATATATGACAGAGCGGCTCTGCAATTTGTACAAAGTACACCGCTTTTAAATCC 1735
Qy SerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIle 492
Db CAGGCTACGATTCAGATGAGCCCTTCCAGCGCGGGAAGATTAATTAATGCTCCATT 1815
Qy LysGlnGlnIleAlaIleThrSerGlyGlnTyrGlnValLeuGlyArgHisGlySerAsn 512
Db AAGGAAAGATTCCTGACAGCGGCTGAAATGGAGGTTTGGGAGGACCGGCTCC--- 1872
Qy IleGlnValAspGluValArgArgLeuValTyrPheGlnIleThrLysAspSerProLeu 532
Db -----AAGGACACAGAGACCGCCCTG 1896
Qy GluHisIleLeuTyrValAlaSerTyrValAsnProGlyGluValThrArgLeuThrAsp 552
Db GAGCAGCAGCTTACGTCGTCAGCTATGAGCGCGCGGAGATTCGATCCGCTCAGCAGC 1956
Qy ArgGlyTyrSerHisSerCysHisSerGlnHisCysAspPhePheIleSerLysTyr 572
Db CCGGCTTCTCCATGACTCTCATGAGCAAACTTCACATGTTCTGTCAGCACCTAC 2016
Qy SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 592
Db AGCAGCTGACAGCGCGCTCTGCTGTCACGTCATCAAGCTTGAAGCGCCCGCAGAGAC 2076
Qy ProThrCysLysThrLysGlnPheTyrPheAlaThrIleLeuAspSerAlaGlyProLeuPro 612
Db CCCCTGCAAGAGCGCCCGCTCTGGGCTAGCATATGAGAGGACAGCTGCGCCCGC 2136
Qy AspTyrThrProProGlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632
Db GATTATGTTCTCCAGAGATCTTCATTTCCACACCGCGCTCGATGTGCGCTCTACGAGC 2196
Qy MetLeuTyrIleProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652
Db ATGATCTCAAGAGCCCAAGCGCTTGCAGCAGCAGGAAAGAACCCCAAGCTCTCTTGTGA 2256
Qy TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 672
Db TATGAGAGCGCCCAAGGTGAGCTGTGATTAATCTCTTCAAGAGCATCAAGTACTGGCG 2316
Qy LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgLysSerCys 692
Db CTCAACACATGCGCTCTCTGCGGCTACCGCGGTGTGATTTACGCGCAGAGGCTCTCTGT 2376
Qy HisArgGlyLeuLysPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 712
Db CAGCAGAGGCGCTGCTGTCAGAGGCGCTGAAACCAATATGGCCAGGTGAGATCGAG 2436
Qy AspGlnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 732

Db 2437 GACCAAGTGAAGGCTCGCATTCGTGGCCGAGAAGTAGGGCTTCATCGACCTGAGCCGA 2496
Qy 733 VALGLYLEHISGLYTRPSEYTRGGLYTYRLEUSELEUMETALAEUMETGLNARG 752
Db 2497 GTTGCATTCATGCGTGTCTTAACGGGGCTTCCTCGCTCATGGGGCTTAATCCACAAAG 2556
Qy 753 SerAsp1LeuPheArgValAlaIleAlaIleValAlaProValThreLeuTrpIlePheTyrAsp 772
Db 2557 CCCCAAGTGTTCAAAGGTGGCCATCGCGGGTGGCCCGGTCAACGTCGTGATGGCTACGAC 2616
Qy 773 ThrGlyValThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeu 792
Db 2617 AAGGGGTAACTGAGGCTTAACAAGACGTCCCTGAAGAACCAAGACAGCGCTATGAGGCG 2676
Qy 793 GlySerValAlaMetClnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 812
Db 2677 GGTTCGTGGCTCGACGTGGAGGAAAGCTGCCCAATGAGCCCAACGGCTTCATTATCTTC 2736
Qy 813 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
Db 2737 CACGGCTTCCTGGACGAAACGTCACCTTTTCCACAAACTTCCTCGCTCCCAACTG 2796
Qy 833 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852
Db 2797 ATCCGAGCAGGAGAACTTAACAGCTCCAGATCTACCCCAACGAGACACAGTATTCCG 2856
Qy 853 ValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 872
Db 2857 TCCCCGAGCTCGGGCGACGACATGAAAGTACCTTCGCTGCTCACTTCTCAAGAAATACCTC 2916
RESULT 14
US-09-976-674-34
; Sequence 34, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
; FILE REFERENCE: 70669
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 4263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-34
Alignment Scores:
Pred. No.: 0 Length: 4263
Score: 2820.50 Matches: 510
Percent Similarity: 76.43% Conservative: 132
Best Local Similarity: 60.71% Mismatches: 183
Query Match: 60.01% Indels: 15
DB: 10 Gaps: 3
US-10-070-464-1 (1-882) x US-09-976-674-34 (1-4263)
Qy 35 PheTyrValGluArgTyrSerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLys 54
Db 436 TTCAGAGTGCAGAACACTCGTGGAGCGGCTCCGAGATATCATCAGCGCAGCGCAG 495
Qy 55 TyrHisGlyTyrMetClnAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74
Db 496 TACTCGGGCTCATGTGTAACAAAGCGCCCGCCAGCTTCAAGTTGTGCAGAAAGCGAGT 555
Qy 75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGlu 94

Db 556 GAGTCTGGGCCCCACTCCACCGCTCTACTACTGCGAATGCCAATATGGCAGCCGAGAG 615
Qy 95 AsnThrLeuPheTyrSerGluIlePheLysValThrIleAsnArgAlaValAlaLeuMetLeu 114
Db 616 AACTCCCTCTACTACTGATATCCCAAGAAAGTCCGGAAGAGCTCTGCTGCTCTG 675
Qy 115 SerTrpLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
Db 676 TCTGTAAACAGATGTGTGATCATTTCCAGGCCACCCCAACATGGAGCTACTCTCGG 735
Qy 135 GluGluGluLeuLeuArgLysArgLysArgLysValGlyValAlaIleAlaSerTyrAsp 154
Db 736 GAGAGAGACTGCTGAGGAGCGAAACCTGGGGGTCTTCGGCATACCTCTCAAGAC 795
Qy 155 TyrHisGlnGlySerGlyTyrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174
Db 796 TTCACAGGAGGAGTGGCTCTTCCTTCACGGCCAGCAACGCTTCCTCCACCTCGC 855
Qy 175 AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194
Db 856 GACGGGGCAAGAACGCTTCATGTGTCTCCCTATATAACCGTGGAAATCAAGACCCAG 915
Qy 195 CyProAsnIleArgMetAspProLysLeuCyProAlaAspProAspTrpIleAlaPhe 214
Db 916 TCTCAGAGCCCGGATGAGCCCAAAATCTGCGCCGACCTGCTCTCTCTCCTC 975
Qy 215 IleHisSerAsnAspIleTrpIleSerAlaIleValThrArgGluGluArgGluLeuThr 234
Db 976 ATCAATAAACAGGACTGTGGGTGGCAATCTGACACAGGAGGAGCGCGCTGAGCC 1035
Qy 235 TyrValHisAsnGluLeuLeuAlaAsnMetGluGluAspAlaArgSerAlaValAlaThr 254
Db 1036 TTCTGCCACCAAGTTTATCCATATGCTCCGAGTAGACCCCAAGTCTCGAGTGGCCAC 1095
Qy 255 PheValLeuGlnGluGluPheAspArgTyrSerGlyTyrTrpTyrCyProLysAlaGlu 274
Db 1096 TTCGTCAATCAGAAAGAGTTGACCGCTTCATCGGGTACTGGTGGGCCCAAGCTCC 1155
Qy 275 ThrThrProSerGlyGlyLysIleLeuValGlyIleLeuTyrGluGluAsnAspGluSer 293
Db 1156 TGGGAAGTTCCAGAGGCTCTCAAGACGCTGCAMATCTGTATGAGAGTCAATGATGTC 1215
Qy 294 GluValGluIleIleValThrSerProMetLeuGluThrArgArgAlaAspSerPhe 313
Db 1216 GAGGTGAGGTATTCAGTCCCTCTCTGCGCTTGAAGAAAGAAAGCAAGCATCTGAT 1275
Qy 314 ArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGluLeuMet 333
Db 1276 CGGTACCCGAGACAGGACAGCAAGATCCCAAGATTGCTGAAACTGAGTCCAGTCCAG 1335
Qy 334 IleAspAlaGluGlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGlu 353
Db 1336 ACTGACACGACGAGGAGATGCTCTCGAACCCAGAGAGAGACTGTCGAGCCCTTCAC 1395
Qy 354 IleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyr 373
Db 1396 TCGCTGTTCCGAAAGTGAAGTACATCGCCAGCGCCGGGTGAACCCGGGATGCAAAATTC 1455
Qy 374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393
Db 1456 GCTTGAGCCATGTTCTGAGACGCGCCAGCAGTGGCTCCAGCTGCTCTCCCTCCCGG 1515
Qy 394 GluLeuPheIleProValGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413
Db 1516 GCCCTGTTCATCCGAGACAGAGATAGAGACGCGGTAGCCTTCGCAAGACTGTC 1575
Qy 414 ProAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTrpIleAsnIle 433
Db 1576 CCCAGGAATGTCCAGCCCTATGTGTGTAACAAGAGTCAACCAAGCTCTGATCAATGTT 1635
Qy 434 HisAspIlePheHisValPheProGlnSerHisGluGluGluIleGluPheIlePhe 452
Db 1636 CATGACATCTTATCCCTCCCAATCAGAGGAGAGAGAGCTGCTGCTTCTCCGC 1695

[illegible]

QY	813	HisGlyPheLeuAspGluAsnValHisPheAlaHisIsthSerIleLeuLeuSerPheLeu	855
Db	2737	CAGGGCTCTCGACGAAAAAGCTGACCTTTTCCACAAACCTTCCTGCTCCCACTG	273
QY	833	ValArgAlaGlySerProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg	855
Db	2797	ATCCGACGAGGAAACCTTACCAAGCTCCAGATCTACCCCAAGAGACACAGATTCCG	285
QY	853	ValProGluSerGlyGlnHisTyrGluLeuHisIleLeuHisTyrLeuGlnGluAsnLeu	872
Db	2857	TGCCCGAGTCGGGCGAGCAGCATATGAGTCAAGTTCGTGTCACCTTCTACAGAAATACCTC	291
RESULT 15			
US-09-976-674-32			
; Sequence 32, Application US/09976674			
; Patent No. US20020115843A1			
; GENERAL INFORMATION:			
; APPLICANT: Qi, Steve			
; APPLICANT: Aktinanya, Karen			
; APPLICANT: Riviere, Pierie			
; APPLICANT: Junien, Jean-Louis			
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV			
; FILE REFERENCE: 70669			
; CURRENT APPLICATION NUMBER: US/09/976,674			
; CURRENT FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: US 607240.117			
; PRIOR FILING DATE: 2000-10-12			
; NUMBER OF SEQ ID NOS: 61			
; SOFTWARE: Patentin version 3.1			
; SEQ ID NO 32			
; LENGTH: 4076			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-976-674-32			
Alignment Scores:			
Pred. No.: 7,06e-309 Length: 4076			
Score: 2649.00 Matches: 486			
Percent Similarity: 72.89% Conservative: 127			
Best Local Similarity: 57.79% Mismatches: 177			
Query Match: 56.36% Indels: 51			
DB: 10 Gaps: 3			
US-10-070-464-1 (1-882) x US-09-976-674-32 (1-4076)			
QY	35	PheTyrValGluArgTyrSerTyrSerGlnLeuGlyIleuValAspThrArgLys	54
Db	436	TTCAGAGTCGACGACGCTCGTGGAGCGGCTCCGAGACATCTACCGAGCGGCGAG	459
QY	55	TyrHisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValLysArgAsp	74
Db	496	TACTCGGGGCTCATTTGTCACAAAGCGGCGCCACGACCTTCAGTTTGTGGAGAAACCGAT	555
QY	75	ProAspGlyProHisSerSerAspArgLLeTyrTyrLeuAlaMetSerGlyLysAsnArgL	94
Db	556	GAGCTCGGGGCGGACCTCCACCGGCTCTTACTGAGAAATGCGATATGCGAGCGGAG	615
QY	95	AsnThrLeuPheTyrSerGlnIleProLysThrIleAsnArgAlaAlaValLeuMetLeu	114
Db	616	AACCTCCTCTGACTGAGATTCGCCAAGAAAGTCCGAAAGAGCTCTGCTGCTG	675
QY	115	SerTyrLysPheLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg	134
Db	676	TCTCGAAGACAGATGCTGGATCATTTCCAGGCGCACGCCACATGAGGAGTCACTCGG	735
QY	135	GluGluGlnLeuLeuArgGluArgLysArgLLeGlyThrValGlyIleAlaSerTyrAsp	154
Db	736	GAGGAGGAGCTCTGAGGAGGAGAAAGCTTGAGGAGCTTGGGACATCACTCTCGAC	795
QY	155	TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys	174
Db	796	TTCACAGGAGAGTGGCTCTTCTCTTCCAGGCGACGAAAGAGCTTCTCACTGCGG	855

QY 175 AspGlyValProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGlnThrSer 194
 DB 856 GACGGGGGCAAGAACCGCTTCATGCTGTCCCTATGAAACCGCGAATCAAGACCCAG 915
 QY 195 CysProAsn1LeArgMetAspProGlyLeuGlyProAlaAspProAspTrp1LeAlaPhe 214
 DB 916 TGTCTACGGGCCCCGGATGAGACCCCAAAATCTGCTCCGACCTGCTCTCTCTCTCC 975
 QY 215 IleHisSerAsnAsp1LeTrp1LeSerAsn1LeValThrArgGlnGlnValArgLeuThr 234
 DB 976 ATCAATACAGCCACCTGTGGGTGGCCAACTGACAGCAGGCGAGGCGGCGGTGAC 1035
 QY 235 TyrValHisAsnGlnLeuAlaAsnMetGlnGlnAlaAsp1LeArgSerAlaGlyValAlaThr 254
 DB 1036 TTCTGCACACCAAGTTTATCAATGTCTCGATGACCCCAAGTCTGGGGGTGGCCAC 1095
 QY 255 PheValLeuGlnGlnGlnGlnPheAspArgTyrSerGlyTyrTrpTrpCysProGlyAlaGln 274
 DB 1096 TTGTCATACAGAGAGGTTCGACCGCTTCACTGGGTACGTGGTCCCCACAGCTCC 1155
 QY 275 ThrThrProSerGlyGly--LysIleLeuArgIleLeuTyrGlnGlnGlnAsnAspGlnSer 293
 DB 1156 TGGGAAGGTTCAAGGGCTTCAAGACGCTCGAATCTGTATAGAGAGTCATGAGTCC 1215
 QY 294 GlnValGlnIleIleHisValThrSerProMetLeuGlnThrArgAlaAspSerPhe 313
 DB 1216 GAGGTGAGGTCAATTCACGTCCCTCTCTGCGCTAGAAAGAAAGAACGACGCTGAT 1275
 QY 314 ArgTyrProGlyThrGlyThrAlaAsnProGlyValThrPheGlyMetSerGlnIleMet 333
 DB 1276 CGGTACCCCAAGCAGCAGCAGCAAGAAATCCCAAGATGCTTGAATCGCTGAGTTCAG 1335
 QY 334 IleAspAlaGlnGlnGlnArgIleIleAspValIleAspGlyLeuLeuGlnProPheGln 353
 DB 1336 ACTGACAGCCAGGCGCAAGATCTCTGACCCAGAGAGAGAGCTGTGCGACCTTCAC 1395
 QY 354 IleLeuPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 373
 DB 1396 TCGCTGTTCCCGAAGGTGGAGTATCATGCGCAGGGCGGGGTGACCCGAGTGGCAATAC 1455
 QY 374 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393
 DB 1456 GCGTGGCCATGTTCTCTGACCGGCCCCCAGCAGTGTCTCAGCTGCTCTCTCTCTCC 1515
 QY 394 GlnLeuPheIleProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerVal 413
 DB 1516 GCCCTGTTATCCCGAGCAGCAGAGATGAGAGCAGCGGCTACCTCTGCCAAGCTGTC 1575
 QY 414 ProAspSerValThrProLeuIleIleTyrGlnGlnGlnThrThrAsp1LeTrp1LeAsnIle 433
 DB 1576 CCCAGGAATGTCACGCGTATGTGTATGAGAGAGGTCAACAAGTCTGATCAATGTT 1635
 QY 434 HisAsp1LePheHisValPheProGlnSerHis---GlnGlnGlnIleGlnPheIlePhe 452
 DB 1636 CATGACATCTTCATCC 1695
 QY 453 AlaSerGlnCysArgThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuValGln 472
 DB 1696 GCCAATGAATGCAAGACCGGCTCTGCACTTGTACAAAGTCAACCGCGTTTAAATCC 1755
 QY 473 SerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIle 492
 DB 1756 CAGGGCTACGATGAGTGGAGGCTCTCAGCCCGGGGAGAGATGAATTTAAGTCCCAT 1815
 QY 493 LysGlnGlnIleLeuIleThrSerGlyGlnTyrGlnGlnValLeuGlnIleArgHisLeuAsn 512
 DB 1816 AAGGAAGAGATTGCTCTGACACCGGAGTGAATGGAGGTTTGGCGAGGCGACGGCTCAAG 1875
 QY 513 IleGlnValAspGlnValArgArgLeuValTyrPheGlnGlnIleThrLysAspSerProLeu 532
 DB 1876 ATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCGCAAGAGACAGCGCGTG 1935
 QY 533 GlnHisHisLeuTyrValValSerTyrValAsnProGlyGlnValThrArgLeuThrAsp 552

DB 1936 GACACACACCTTACAGTGGTCACTATGAGCGCGGCGGAGATTCGATACGCTCACCAG 1995
 QY 553 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPheIleSerLysTyr 572
 DB 1996 CCGGCTTCTCCATAGCTGTCTCAATGAGCAGAAACTCAATGTTCGTGACGACTAC 2055
 QY 573 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGlnAspAsp 592
 DB 2056 AGCAGGCTAGACACCGCCCTCGTGCACGTCTTCAAGCTGAGCGGCCCGCAGACGAC 2115
 QY 593 ProThrCysLysThrLysGlnPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 612
 DB 2116 CCCCTGCAACAGACCGCCGCTTCTGCGGTAGATATGAGGACGACGACCTGCCCCCG 2175
 QY 613 AspTyrThrProProGlnIlePheSerPheGlnSerThrThrGlyPheThrLeuTyrGly 632
 DB 2176 GATTATGTTCTTCCAAAGATCTTCAATTCACACGCGCGATGTCGCGCTCTACGCG 2235
 QY 633 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652
 DB 2236 ATGATCTCAAGCCCAAGCGCTTGCAGCCAGGAGAGAGACCCACGCTCTTTGTA 2295
 QY 653 TyrGlyValProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 672
 DB 2296 TATGAGAGGCCCCAGGTGACGTGTGAATTAATCTCTTCAAGAGCATCAAGTACTTGG 2355
 QY 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCys 692
 DB 2356 CTCACACACTGGCTCCCTCGGAGTACCGCTGTGTGTATTAACGACGAGGGCTCTCT 2415
 QY 693 HisArgGlyLeuLysPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 712
 DB 2416 CAGCGAGGCTTCTGTTTCAAGGGGCTTGAAGAAACCAATGAGGCGAGGTGAGATGAG 2475
 QY 713 AspGlnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 732
 DB 2476 GACCAAGTGAAGGCTTGCAGTTCGTGGCCGAGAGATAGTGGCTTCACTGACGAGCGGA 2535
 QY 733 ValGlyIleHisGlyTyrPheTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
 DB 2536 GTTGCCATCATGCTGTGTCTTACGAGGGGCTTCTCTGCTCATGGGGCTTAATCCACAG 2595
 QY 753 SerAsp1LePheArgValAlaIleAlaIleAlaGlyAlaProValThrLeuTrp1LePheTyrAs 772
 DB 2596 CCCAGGTTTCAAGG----- 2611
 QY 772 PThrGlyTyrThrGlnArgTyrMetGlnHisProAspGlnAsnGlnGlnGlnGlnGlnGln 792
 DB 2611 ----- 2611
 QY 792 uGlySerValAlaMetGlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeu 812
 DB 2612 -----CCCAACCGCTTGTCTTACCT 2631
 QY 812 uHisGlyPheLeuAspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheIle 832
 DB 2632 CACGGCTTCTGAGAGAAACGTGCACCTTTTCCACAAACTTCTCTCTCCCAACT 2691
 QY 832 uValAlaArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArg 852
 DB 2692 GATCCGAGCAGGGAACCTTACACGCTCCAGATCTACCCCAACAGAGACAGATTTGCG 2751
 QY 852 uValProGlnSerGlyGlnHisTyrGlnLeuHisIleLeuLeuHisTyrLeuGlnGlnLeu 872
 DB 2752 CTGCCCGAGTGGGGGAGACACTATGAACTACGTTGTCGACTTCTACAGAAATACCT 2811
 QY 872 u 872
 DB 2812 C 2812

Job time : 792.423 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 5299 Seconds

(without alignments)
4045.398 Million cell updates/sec

SUMMARIES

29: gb_gss2.*
Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAMETQQLGVEIFETADC.....HLHYLGNTGSRALAKVI 882
Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO-seq/US10070464/runat_15102003_113553_24829/app_query.fasta.1.2652
-DB=EST -QFMT=fastp -SUFFIX=1rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human4.0.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pflo -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464 @CGN 1.1 8056 @runat 15102003 113553 24829 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: gb_gss1.*

Result No.	Score	Query Match	Length	DB ID	Description
1	4528.5	96.4	3143	11	AK016546 Mus muscu
2	4403	93.7	5517	11	AK029788 Mus muscu
3	2833	60.3	3327	11	AK050021 Mus muscu
4	2823	60.1	3457	11	AK078301 Mus muscu
5	2820	60.0	3376	11	AK050023 Mus muscu
6	1693.5	36.0	1042	12	BM557438 AGENCOURT
7	1542	32.8	957	13	BO675006 AGENCOURT
8	1539.5	32.8	910	13	BO675260 AGENCOURT
9	1505.5	32.0	1041	12	BI084090 AGENCOURT
10	1430	30.4	968	13	BO671635 AGENCOURT
11	1427	30.4	910	13	BO671276 AGENCOURT
12	1397	29.7	919	13	BM390898 AGENCOURT
13	1370.5	29.2	853	12	BI223892 AGENCOURT
14	1354	28.8	1082	13	BU239476 AGENCOURT
15	1343	28.6	753	9	AL040398 DME2434
16	1342	28.6	789	10	BM709118 AGENCOURT
17	1342	28.6	3726	11	AK039652 Mus muscu
18	1325	28.2	1090	12	BM910838 AGENCOURT
19	1286	27.4	746	14	BY751026 AGENCOURT
20	1277.5	27.2	985	13	BO068650 AGENCOURT
21	1276	27.1	831	12	BM974324 AGENCOURT
22	1260	26.8	802	14	CD354396 AGENCOURT
23	1248	26.6	735	13	BU447277 AGENCOURT
24	1242.5	26.4	788	10	BM6163397 AGENCOURT
25	1242.5	26.4	1021	12	BM974587 AGENCOURT
26	1225	26.1	741	14	CB595102 AGENCOURT
27	1216	25.9	851	12	BI084885 AGENCOURT
28	1210.5	25.8	993	10	BM259714 AGENCOURT
29	1201	25.6	762	14	CD352521 AGENCOURT
30	1196	25.4	932	14	CA790315 AGENCOURT
31	1192	25.4	695	10	BM715866 AGENCOURT
32	1171	24.9	910	10	BM479035 AGENCOURT
33	1170	24.9	897	10	BM784154 AGENCOURT
34	1136	24.2	897	13	BU191638 AGENCOURT
35	1130.5	24.1	674	13	BU229303 AGENCOURT
36	1127.5	24.0	774	14	CB231750 AGENCOURT
37	1127	23.8	669	10	BM390103 AGENCOURT
38	1120	23.8	1002	10	BM733691 AGENCOURT
39	1117	23.8	674	10	BM888665 AGENCOURT
40	1099	23.4	639	10	BM165205 AGENCOURT
41	1081	23.0	658	10	BM388695 AGENCOURT
42	1080.5	23.0	853	10	BM694553 AGENCOURT
43	1073	22.8	661	10	BM799580 AGENCOURT
44	1066	22.7	617	10	BM789604 AGENCOURT
45	1066	22.7	845	13	BU216915 AGENCOURT

ALIGNMENTS

RESULT 1
AK016546
LOCUS
DEFINITION
AK016546 3143 bp mRNA linear HTC 05-DEC-2002
MUS musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog
[Homo sapiens], full insert sequence.
ACCESSION
AK016546
VERSION
AK016546.1 GI:12855334
KEYWORDS
HTC, CAP trapper
SOURCE
MUS musculus (house mouse)
ORGANISM
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1	Carninci, P. and Hayashizaki, Y.
AUTHORS		
TITLE		High-efficiency full-length cDNA cloning
JOURNAL		Mech. Enzymol. 303, 19-44 (1999)
PUBMED		99279253
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
AUTHORS		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL		prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED		Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	3	
AUTHORS		Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,
		Kono, H., Akiyama, J., Nishi, K., Katsuna, T., Tashiro, H., Itoh, M.,
		Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
		Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
		Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
		Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL		sequencing pipeline with 384 multicapillary sequencer
PUBMED		Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE	4	
AUTHORS		Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
		Arkawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
		Aizawa, K., Iwata, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,
		Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R.,
		Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
		Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H.,
		Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, T., Peele, G.,
		Quackenbush, J., Schriml, L.M., Staahl, F., Suzuki, R., Tomita, M.,
		Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
		Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
		Carninci, P., de Bonaldo, M.F., Brownstein, M., J., Bult, C.,
		Fletcher, C., Fujita, M., Gaidboldi, M., Gustincich, S., Hill, D.,
		Hotchkorn, M., Hume, D. A., Kamuya, M., Lee, N.H., Lyons, P.,
		Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P.,
		Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
		Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
		Toyo-Oka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L.,
		Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohzuki, S.
		and Hayashizaki, Y.
TITLE		Functional annotation of a full-length mouse cDNA collection
JOURNAL		Nature 409 (6821), 685-690 (2001)
PUBMED		21085660
REFERENCE	5	
AUTHORS		The FANTOM Consortium and the RIKEN Genome Exploration Research
TITLE		Group Phase I & II Team.
JOURNAL		Analysis of the mouse transcriptome based on functional annotation
PUBMED		of 60,770 full-length cDNAs
REFERENCE		Nature 420, 563-573 (2002)
AUTHORS		6 (bases 1 to 3143)
		Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
		Arkawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
		Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
		Hishio, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
		Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Konda, M.,
		Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
		Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
		Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
		Shibata, K., Shibata, Y., Shinagawa, A., Shikata, T., Sogabe, Y.,
		Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
		Tejima, Y., Toyama, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
		Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of
		Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT

EXPLORATION Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAAGCATTCGAAGCCTCTCTTTTCTTTTTTNN 3'], cDNA was
prepared by using trihalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GGAGAGCAATTCGCAGTTAATTAAATTATCCCCCCTCCCC 3']. cDNA was cleaved
with BamHI and XhoI. cDNA of size comprised longer than 7 kb was
selected before cloning. Vector: a modified plasmidscript KS(+) after
bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3'
end: BamHI. Host: DH10B.

FEATURES

source

location=Qualifiers

1..3143

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:4932434F09"

/db_xref="MG1:1907554"

/db_xref="taxon:10090"

/clone="4932434F09"

/sex="male"

/tissue_type="testis"

/clone_id="Riken full-length enriched mouse CDNA library"

/dev_stage="adult"

327..3005

/note="unnamed protein product: DIPERTIDYL PEPTIDASE 8
homolog [Homo sapiens] (SPRR109HM5, evidence: FASTY,
95.%ID, 100%length, match=2649)
putative"

/codon_start=1

/protein_id="BAB30295.2"

/db_xref="GI:26382128"

/cd_xref="MG1:1921638"

/translation="KKIPSGRCMAAMATEQUGVEIFETAECEEGNSGSDRPKLEP
FYVERYSWSQLKLADVTRKHGYMMAKAHDPMFAVRTDPDRPHSRVYLLAMSGEN
RENTLFPSELPKTINRAAVLMLSMKPLDLDFQATLYGMASREELKERKRGTVGI
AAVDYHPGKTFLEQNGSIGYIH KDGPHEFTOOPRLPNVENSCHIRMDPKCPAD
PDMAIFHSNDIWINSLVPEERRITYNHLELMNEDPSGAVTAFTVLDOEPFGYSG
YMKCPOAPRPDSGCKILRTLYENDSSGVIIHYTSMTETRADASRYKCRANK
VTKMSIEIVDAAGGITIDVIDKELVQFFELFEVEVETIAAGATTPEGKAMSTILDS
QTHLQVLIPELFIPVEDAMDQRRLIESVPSTPLLIEETDITWINIHDIFHFV
PORHEDIELFIASEECTGTFRHLVKITS LKESKYKSSGGLAPASPFCPIKEITI
TSGEMELGHGRGINIWDEARKLYVEGTSDPLEHLVYTSANPEVRLIDRGYS
HSCSLSHCDPFISKYSNOGNPCVS LYKLSPEDDPVHKTEFMATILDSAGLPGY
TPLEISFESTTGFTLYGMYKPXHDLOPKKPYTLFIYGPOVOLVNPRFKVYKR
LNLIASGYVVVINDRGSCHRLKKBGAKGQIIEIDQVEGLDYLAQDFIDL
DRYIGHMSYGYLSLALMORSDIREFVAIAGAPVLTMYDTDGTERBTGHPDQSO
GYLGVSVMQAERKFPSEPNRLILLHGFLDENVFHTSIJLSLTVRKGQYDVQIYPQ
ERSISRVBEGEYEHLTHLYLDENIGSRIALTKVI"

BAB30295.2

BASE COUNT

894 a 688 c 773 g 768 t

ALIGNMENT

Alignment Scores:

Pred. NO.: 0

Score: 4528.50

Percent Similarity: 97.73%

Best Local Similarity: 95.70%

Query Match: 96.35%

Length: 3143

Matches: 845

Conservative: 18

Mismatches: 19

Indels: 1

Gaps: 1

US-10-070-464-1 (1-882) x AK016546 (1-3143)

1 MecAlaAlaLaMetcgturthrnguglnleuGluValglutlIePheglturhralaaapCyS 20

|||||
Db 354 ATGGAGCAGCAGCATGGAAACAGACAGCTGGGTGAGATATTTGAACTGCAGAGTGT 413
|||
Qy 21 GUGUGU---Asn11Eg1uSerG1naSpaPrOlvSLeuG1uPProPheTYrValG1uaG 39
|||
Db 414 GAGGAGGGAATGCGAATCAGAGACCAACCACTGAGCCCTTTATATGCGAAGCG 473
|||
Qy 40 TYrSerTPSerG1nLeuYs1ySLeuLeuA1aAspThArG1yS1yS1yTYrMet 59
|||
Db 474 TATTTCTGGAGCTCAGCTGAAAAAGCTGCTTCTGATACCGAAAAATACAGCGCTCAGT 533
|||
Qy 60 MetAlaYsAlaProH1saPheMetPheValYsArgAsnAspProAspG1yProH1s 79
|||
Db 534 ATGGCTAAGCGCCCATGATGATTTATGTTGTAAGAGAGCGATCCAGATAGGCTCAG 593
|||
Qy 80 SerAspArg11eTYrTYrLeuAlaMetSerG1yG1uAsnArgG1uAsnThrLeuPheTYr 99
|||
Db 594 TCAGACAGGCTGATATTACTTGCATGTGAGAAACAGAGAAATACACTATTTTAT 653
|||
Qy 100 SerG1u11ePro1yS1yThr11eAsnArgAla1aValLeuMetLeuSerTP1ySPro1eu 119
|||
Db 654 TCTGAATCTCTAAACCATCAACAGACAGCAGCTTTATGCTTTCTCGAAGCCCTT 713
|||
Qy 120 LeuAspLeuPheG1na1aThrLeuAspTYrG1yMetTYrSerArgG1uG1uLeuLeu 139
|||
Db 714 TTGGATCTTTTTCAGGCGACACTAGACTATGGGATGTATTTCTGAGAGGAAAGACTATC 773
|||
Qy 140 ArgG1uaG1ySArgG11eG1yThrValG1y11eAlaSerTYrAspTYrH1sG1nG1ySer 159
|||
Db 774 AGAGAAAGAAAGCGCATTTGGAACCGTGCGGATCGCAGCTTATGATTAACACACCGAAGAC 833
|||
Qy 160 G1yThrPheLeuPheG1na1aG1ySerG1y11eTYrH1sVal11yAspG1yG1yProG1n 179
|||
Db 834 GGAATCTCTCTTTCAAGCTGCGACGCGCATTTACATATCAAGATGAGAGACCAACT 893
|||
Qy 180 G1yPheThrG1nG1nPro1euArgProAsnLeuValG1uThrSerCySProAsn11eArg 199
|||
Db 894 GGATTTTACAAACACCTTTCGCGCCCAATTAGTGAATTAAGTTGTCTCCATATACGA 953
|||
Qy 200 MetAspPro1ySLeuCySPro1a1aAspProAspTP11eAlaPhe11eH1sSerAsnAsp 219
|||
Db 954 ATGGATCCAAATATATGCTCTGATCCATGACAGCTGATAGCTTTTCACTCAGCAGCATGAT 1013
|||
Qy 220 11eTP11eSerAsn11eValThrArgG1uG1uaArgLeuThrTYrValH1sAsnG1u 239
|||
Db 1014 ATTGGATATCAACTGCTGTAACAGGAGAGCGAGAGATCACTACTACCAATGAG 1073
|||
Qy 240 LeuAlaAsnMetG1uG1uAspAlaArgSerAlaG1yValAlaThrPheVal11eG1nG1u 259
|||
Db 1074 CTAGCCAAACATGGAAGAGGATCCAGATCAGCTGGGTGCGACCTTGTCTTCAGAGA 1133
|||
Qy 260 G1uPheAspArgTYrSerG1yTYrTP1yP1yCySPro1ySAlaG1uThrThrProSerG1y 279
|||
Db 1134 GAATTTGACAGATCTGCTGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1193
|||
Qy 280 G1y1yS11eLeuArg11eLeuTYrG1uG1uAsnAspG1uSerG1yValG1u11eH1s 299
|||
Db 1194 GGTAAATTTCTTAGAATCTCTATGAAGAAATGATGATCTGAGGTGAGATTTATTCAT 1253
|||
Qy 300 ValThrSerProMetLeuG1uThrArgArgAlaAspSerPheArgTYrPro1ySThrG1y 319
|||
Db 1254 GTTACGTCTCCCATGTTGGAACCAAGAGGCGAGATTCTTCTGTTATCCCAAAACAGGC 1313
|||
Qy 320 ThrAlaAsnPro1ySValThrPhe1ySmetSerG1u11eMet11eAspAlaG1uG1yArg 339
|||
Db 1314 ACAGCAAAACCAAGAGTCACTTCAAGAGTGGAGATGTTGTTGATGCTGCAGAGAGCG 1373
|||
Qy 340 11e11eAspVal11eAsp1ySLeuLeu11eG1nProPheG1u11eLeuPheG1uG1yVal 359
|||
Db 1374 ATTATAGATGCTATGATTAAGAACTGGTCAACTTTGAGATCTGTTTGAAGGAGAT 1433
|||
Qy 360 G1uTYr11eAlaArgAlaG1yTYrThrProG1uG1y1yS1yTYrAlaTP1ySer11eLeuLeu 379
|||

|||||
Db 1434 GAATATATTCCAGAGCCGAGATGCACTCCAGAGGAAAAACATGCTGTCACTACTA 1493
|||
Qy 380 AspArgSerG1nThrArgLeuG1n11eVal11eSerProG1uLeuPhe11eProVal 399
|||
Db 1494 GACCGTTCCCAACTCACTCGAGATAGTTGATCTCTCCGAGTATTCATCCAGTA 1553
|||
Qy 400 G1uAspAspVal11eG1uArgG1nArgLeu11eG1uSerValProAspSerValThrPro 419
|||
Db 1554 GAAGATGATGCCATGACAGACAGACACTTAATAGACTAGTTCTGACTCTGAGACCA 1613
|||
Qy 420 Leu11e11eTYrG1uG1uThrThrAsp11eTP11eAsn11eH1sAsp11ePheH1sVal 439
|||
Db 1614 CTGATCATCTATGAAGAAACACAGACATCTGATTAATATCCAGATATTTTCATGTT 1673
|||
Qy 440 PheProG1nSerH1sG1uG1uG1u11eG1uPhe11ePheAlaSerG1uCyS1ySThrG1y 459
|||
Db 1674 TTTCTCAAACTCAATGAAGATGAATAATGATTTATTTTCTGTAATGCAAAACAGGT 1733
|||
Qy 460 PheArgH1sLeuTYr11eThrSer11eLeu1yS1ySerTYr1ySArgSerSer 479
|||
Db 1734 TTTCTCATCTGATTAATTCATCTCATTTTAAAGAGACCAATATTAACGCTCACT 1793
|||
Qy 480 G1yG1yLeuProAlaProSerAspPhe1yS1yPro11e1ySLeuG1u11eAla11eThr 499
|||
Db 1794 GGTGAGTACCTGCCCCAAGTATTTCAAGTGTCTCATCAAGAAAGAAATTAACAATTACC 1853
|||
Qy 500 SerG1yG1uTP1yG1uVal11eG1yArgH1sG1ySerAsn11eG1nValAspG1uValArg 519
|||
Db 1854 AGTGTGTAATGGAAAGTACTTGGCCGCGATGATTAATCTGGTTGATGTAAGCCAA 1913
|||
Qy 520 ArgLeuVal11yTYrPheG1uG1yThr1ySAspSerPro1euG1uH1sH1sLeuTYrVal 539
|||
Db 1914 AAGCTGGTACTTTGAAAGCACCAAGACTCTCTTGGACATCACCTTGACGTGACC 1973
|||
Qy 540 SerTYrValAsnProG1yG1uValThrArgLeuThrAspArgG1yTYrSerH1sSerCyS 559
|||
Db 1974 AGTTATGCAAACTCGAGAAAGTGGAGGCTGACGACCGTGGCTACTACACTCTCCTG 2033
|||
Qy 560 CyS11eSerG1nH1sCySAspPhe11eSer1yTYrSerAsnG1n1ySAsnProH1s 579
|||
Db 2034 TGCCCTCACCGGCATTTGCACTTCTTCATAGTAAGTACAGCAACAGAAATCCACAC 2093
|||
Qy 580 CySVal11eLeuTYr1ySLeuSerSerProG1uAspAspProThrCyS1ySThr1yS1yG1u 599
|||
Db 2094 TGTGTCTCTTACAAACTCTCAAGCTCTGAGATGACCCGTTCAATTAACAAAGGAA 2153
|||
Qy 600 PheTP1a1aThr11eLeuAspSerAlaG1yProLeuProAspTYrThrProProG1u11e 619
|||
Db 2154 TTTTGCGCCACCATTTTGATTCACAGGTCCTCTTCGACTACACCCCTCAGAAATT 2213
|||
Qy 620 PheSerPheG1uSerThrThrG1yPheThrLeuTYrG1yMetLeuTYr1ySProH1sAsp 639
|||
Db 2214 TTTTCTTTTGAAGATCACTGATTTACAGTATGGAATGTGTATTAAGCTCATGAC 2273
|||
Qy 640 LeuG1nProG1y1ySLeuTYrProThrVal11ePhe11eTYrG1yG1yProG1nValG1n 659
|||
Db 2274 CTACAACTCGAAAGAAATTAACCCACTGTGTTATTCATATATAGGAGTCCCAAGGCGAG 2333
|||
Qy 660 LeuVal1aAsnAspArgPhe1yS1yVal1yS1yPheArgLeuAsnThrLeuAlaSerLeu 679
|||
Db 2334 CTGGTGAACCAATCGGTTTAAAGAGTCAAGATTTCCCGCTGAAACACCTGCGCTCCCG 2393
|||
Qy 680 G1yTYrVal1aVal1a11eAspAsnArgG1ySerCyS1ySArgG1yLeu1ySPhuG1u 699
|||
Db 2394 GGTATATGTTGTGTATAGCAACAGGGGATCTCTGACCGAGGACTTAATTTTGA 2453
|||
Qy 700 G1yAlaPhe1yS1yTYr1ySmetG1yG1n11eG1u11eAspAspG1nValG1uG1yLeuG1n 719
|||
Db 2454 GCGCGCTTAAATATTAATGGGTCAATTAAGATCAATCAAGTCAAGTGAAGGAGCTCCAG 2513
|||
Qy 720 TYrLeuAlaSerArgTYrAspPhe11eAspLeuAspArgValG1y11eH1sG1yTP1ySer 739
|||
Db 2514 TACCTAGATCTCAGTATGATGATTCATTGACTTGGATGCAAGTGGGATCCAGGCGGTGCC 2573
|||

OY	740	TyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAspSerAspIlePheArgValAla	759
Db	2514	TATGGTGGCTACCTCTCCCTGATGCGATTATTCAGAGGTGGGATATCTTCGGGTGCT	2633
OY	760	ILeAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyr	779
Db	2634	ATTCTGGGGGCCCCAGTCACCCCTGGATCTTTCTATGATACAGATACACGGAGGGCTAT	2693
OY	780	MetGlyHisProAspGlnAspGlnGlnIleTyrTyrLeuGlySerValAlaMetGlnAla	799
Db	2694	ATGGGTGACCCCTGACCAATGAAGAGGGCTACTTATAGGATCTGGGGCCATCCAAACCG	2753
OY	800	GluTyrPheProSerGlyProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsn	819
Db	2754	GAGAAAGTCCCCCTCAGAACCAACCCGTTACTCTCTTGCAATGGGTTCTTGGAATGAGAT	2813
OY	820	ValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyr	839
Db	2814	GTTCACTTGGACACACAGCATATATGCTGAGATTATTTAGTAGAGGGCTGAAAGCATAT	2873
OY	840	AspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGlyGluHis	859
Db	2874	GACTTACGAGATCTATCTCCAGAGAGGACAGACATCAAGTTCCTGACTTGGAGAACAC	2933
OY	860	TyrGluLeuHisIleLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeu	879
Db	2934	TATGACATGCACTCTGCTCCTCACTTACAGAGAACCTTGATCCCGTATCCGCTGCTTGT	2993
OY	880	LysValIle 882	
Db	2994	AAAGTGATA 3002	
RESULT 2			
LOCUS	AKO29788	5517 bp	mRNA
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched		
	[Library, clone:4930560C15 product:DIPEPTIDYL PEPTIDASE 8 homolog		
	[Homo sapiens], full insert sequence.		
ACCESSION	AKO29788		
VERSION	AKO29788.1 GI:26081520		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	99279253		
REFERENCE	10349636		
AUTHORS			
	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
		Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
		Normalization and subtraction of cap-trapper-selected cDNAs to	
		prepare full-length cDNA libraries for rapid discovery of new genes	
		Genome Res. 10 (10), 1617-1630 (2000)	
JOURNAL			
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS			
		Shibata, H., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,	
		Kono, H., Akiyama, J., Nishi, K., Kitenau, T., Tashiro, H., Itoh, M.,	
		Sunji, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,	
		Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,	
		Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Wataniki, M.,	
		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,	
		Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	
		RIKEN integrated sequence analysis (RISA) system-384-format	
		sequencing pipeline with 384 multicapillary sequencer	
		Genome Res. 10 (11), 1757-1771 (2000)	
JOURNAL			
MEDLINE	20530913		
PUBMED	11076861		

REFERENCE		Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Alizawa, K., Izawa, M., Nishii, K., Kiyosawa, H., Kondo, S., Yamahaki, I., Saito, T., Okazaki, W., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Cacaevant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Koehnle, H., Knehl, P., Lewis, S., Matsuo, Y., Nikido, I., Pesole, G., Quackenbush, J., Schirral, L.M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldirelli, R., Barsh, G., Blake, J., Boftelli, D., Boujunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Guestinch, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Monbaerts, P., Nordone, P., Ring, B., Rigwald, M., Rodriguez, I., Sakamoto, N., Saeki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyok-Oka, K., Wang, K.H., Welte, C., Whitaker, C., Wilting, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.
TITLE	JOURNAL MEDLINE PUBMED REFERENCE	Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851
AUTHORS		5
TITLE		The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) (bases 1 to 5517)
JOURNAL REFERENCE		Adachi, J., Alizawa, K., Akamura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hamaguchi, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kunitera, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE COMMENT		Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp), URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ location/Qualifiers
FEATURES	SOURCE	1. 5517 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:4930560C15" /db_xref="taxon:10090" /clone="4930560C15" /sex="male" /issue_type="testis" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 2221. 4837 /note="DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens]"
misc_feature		

(SPT|Q9HWS, evidence: FASTY, 95.7%ID, 100%length,
match=2649)
putative"

BASE COUNT 1514 a 1224 c 1230 g 1549 t
ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 5517
Score: 4403.00 Matches: 827
Percent Similarity: 95.48% Conservative: 17
Best Local Similarity: 93.55% Mismatches: 17
Query Match: 93.68% Indels: 23
DB: 11 Gaps: 2

US-10-070-464-1 (1-882) x AK029788 (1-5517)

Qy 1 MetAlaAlaIaMetGluThrGluGluIleuGluValGluIlePheGluThrAlaAspCys 20
Db 2248 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGCGAGATATTTGAAACTGCAAGAGTGT 2307
Qy 21 GluGlu--AsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArg 39
Db 2308 GAGGAGGAAATGCGATCATCAGGACCGACCCAACTGAGACCTTTTATGTGCAACGG 2367
Qy 40 TyrSerTPSerGlnLeuLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMet 59
Db 2368 TATCTTGAGTCAAGCTGAAAGAGCTGCTGCTGATACAGAAATACACAGGCTACATG 2427
Qy 60 MetAlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHis 79
Db 2428 ATGGCTAAGGCGCCACATGACTTATGTTTGAAGAGAGCGATCCGATGGGCTTCAC 2487
Qy 80 SerAspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgLysAsnThrPhePheTyr 99
Db 2488 TCAGACAGGCGTCTATACCTTGCCATGTCTGTGTGGAACAGAGAAATACATATTTTAT 2547
Qy 100 SerGluIleProLysPheThrIleAsnArgAlaAlaValLeuMetLeuSerTyrLysProLeu 119
Db 2548 TCTCAAAATCCCTAAACCATCAACAGACAGACAGCTTAAATGCTTTCCTGGAAGCCCTT 2607
Qy 120 LeuAspLeuPheGlnAlaThrLeuAspPheTyrLysTyrSerArgGluGluLeuLeu 139
Db 2608 TTGGATCTTTTTCAGGCGACACTAGACTAGTGGATGTAATCTTCAGAGGAAGACTCTTC 2667
Qy 140 ArgGluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySer 159
Db 2668 AGAAGAAAGAAAGCCCATTTGGAAACCGTGGGATCCGAGCTTATGATTACACCCAGAAAC 2727
Qy 160 GlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGln 179
Db 2728 GGAACATTCCTGTTCAAGCTGGCAGCGGCAATTACATCAATCAAAAGATGAGAGCACAT 2787
Qy 180 GlyPheThrGlnGlnProLeuAspProAsnLeuValGluThrSerCysProAsnIleArg 199
Db 2788 GGATTTACACACAGACTTGGCGGCCCAATTTAGTGGAACTAGTGTGCCCAAAATATGCA 2847
Qy 200 MetAspProLysLeuCysProAlaAspProAspTyrIleAlaPheIleHisSerAspAsp 219
Db 2848 ATGATCCAAATATATATGCCCTGCGATCCAGACTCGATAGCTTTCATTCACAGCAATGAT 2907
Qy 220 IleTyrIleSerAsnIleValThrArgGluGluArgLysLeuThrTyrValHisAsnGlu 239
Db 2908 ATTTGGTATCAAACTCGTAAACAGGGAAGAGCGGAGATCATCATGTCACATATGAG 2967
Qy 240 LeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 259
Db 2968 CTAGCCAACTAGGAAGAGATCCAGATCAGCTGCTGTGTCACACTTGTTCACAGAA 3027
Qy 260 GluPheAspArgTyrSerGlyTyrTyrProLysAspGluThrThrProSerGly 279
Db 3028 GAATTTGACAGATATCTGTGGTACTGTGGTGTGCCCAAGGGAAGAAATCTCTAGTGGT 3087
Qy 280 GlyLysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHis 299

Db 3088 GGTAAATTTCTAGATTTCTCATGACAAATAATGATGATCTGAGCTGAGATTTATTCAT 3147
Qy 300 ValThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyr-ProLysThrGly 319
Db 3148 GTTACGTCCCACTAGTTGGAAACAGAGAGGCGAGATTCCTTCGTTATCCCAAAA----- 3202
Qy 319 YThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArg 339
Db 3203 ----- 3204
Qy 339 GluIleLeuAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyVal 359
Db 3205 GATTATAGATGTCATAGATTAAGAACTGTTCAACCTTTCGAGATTCGTTTGGAGGAGT 3264
Qy 359 IGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrPheIleLeuLe 379
Db 3265 TGAATATATTCACAGAGCCGATGAGCTCCAGAGGAAACATGCTGCTGCTCATCTACT 3324
Qy 379 uAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProVal 399
Db 3325 AGACCGTCCCAAGACTCACTGCAGATGTTCTGATCTCCCTGAGTTATTCATCCAGT 3384
Qy 399 IGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrPr 419
Db 3385 AGAAGATGATGCTCATGACAGACAGAGACTTATAGACATTTCTGACTCTGTGACACC 3444
Qy 419 GluIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHisVal 439
Db 3445 ACTGATCATCTATAGAAACACAGACATCTGATGATTAATTCACAGATATTTTTCATGT 3504
Qy 439 LProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysPheThrGly 459
Db 3505 TTTTCTCCAACTAGATGAATGAAATTTGAGTTATTTTGGCTCTGATCAACAAACXG 3564
Qy 459 YPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerTyrLysArgSerSe 479
Db 3565 TTTTGCATCTGATATAAATACATCATCTTTTAAAGAGAGCAATTAACGCTCAG 3624
Qy 479 TGLYGLYLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleTh 499
Db 3625 TGGTGACTACTCTCCCAAGTGAATTCAGAGTCTTATCAAAAGAAATTAACATTAAC 3684
Qy 499 TSerGlyGluThrProValLeuGlyArgHisGlySerAsnIleGlnValAspGlyValAr 519
Db 3685 CAGTGTGAATGGGAAGTACTGGCGGCAATGATCTTAATATCTGGGTGATGAAGCCAG 3744
Qy 519 GArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisHisLeuTyrValIva 539
Db 3745 AAAGCTGTGATCTTTGAAGCACCAAGACTCTCTTTGGAGCATCACTGTACGTGAC 3804
Qy 539 LserTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCy 559
Db 3805 CAGTTATGCAAAACCTGAGAAAGTGGAGGCTGACGACGCTGACTACTCACTCCTG 3864
Qy 559 sCyValSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHis 579
Db 3865 CTGCTCAGCGGAGATTTGATCTTTCATAGATTAAGTAAAGCAACCAAGAAATTCACA 3924
Qy 579 sCyValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGly 599
Db 3925 CTGCTGTCTCCTCTACAAACTCTCAAGCTCTGAGAGATGAGCCCACTTCAATAAACAAG 3984
Qy 599 uPheThrPheAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIle 619
Db 3985 ATTTGGGCGCACCATTTTGGATTCAAGAGGTCTCTCTTCTGACTACACCCCTCAGAAAT 4044
Qy 619 ePheSerPheGluSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysProHisAs 639
Db 4045 TTTTCTTTTGAAGATACATGATTTTACATGATGATGATGATGATGATGATGATGATGATG 4104
Qy 639 pLeuGlnProGlyLysLeuTyrProThrValLeuPheIleTyrGlyProGlnValGly 659

Db	4105	CCTACACCTGGAGAAAGAAATATCCCACTGTATTATTCATATATGTGGTGTCCCAAGTGC	4164
Qy	659	nleuVal1asnAsnaAgpPheLyG1Val1yTyrPheArgLeuAenThrLeuAlaSerLe	679
Db	4165	GCTGGTGAACAATCGGTTTAAAGAGATCAAGTATTTCCGCTGAAACCCCTGGCTCCCT	4224
Qy	679	uG1YrVal1Val1Val1Val1LeaSpAsnArgLySerCysh1ArgG1YrLeuLySpheG1	699
Db	4225	GGGTTATGTGGTGTGGATGATGACAAACGGGGATCTCTGTCAACGAGGACTTAAATTTGA	4284
Qy	699	uG1YrVal1PheLyTyrLySerMeG1YrLeu1LeaSpArg1Val1G1uG1YrLeuG1	719
Db	4285	AGGCCCTTTAAATATTAATATGAGTCAATATCAATATCAATGATGATGAGGAGACTCCA	4344
Qy	719	nTyLeu1a1aSer1aG1YrAspPhe1LeaSpLeuAspArgVal1G1YrLeu1Lea1a	739
Db	4345	GTAACCTAGCATTCGTATGATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT	4404
Qy	739	rTyG1YrG1YrLeuSerLeuMetAlaLeuMetG1naArgSerAsp1LeaPheArgVal1A1	759
Db	4405	CTATGGTGGCTACCTCTCCCTGATGGCATTAATGACAGAGTCGATATCTTCGGGTGGC	4464
Qy	759	a1Lea1aG1YrVal1PheVal1ThLeuTrp1LeaPhe1YrAspThrG1YrTrhG1naArgT	779
Db	4465	TATTCCTGGGGCCCAAGTCAACCTGTGGATCTTCTATGATACAGATACACGAGCGCTA	4524
Qy	779	rMeG1Yr1a1aProAspG1naArgG1uG1YrTrp1YrLeuG1YrSerVal1AlaMetG1na1	799
Db	4525	TATGGGTGACCTCTGACAGAAATGACAGGGCTACTTACGATCTGTGGCCATGCAAGC	4584
Qy	799	aG1YrLySpPheProSerG1uProAsnArgLeuLeuLeu1a1aG1YrPheLeuAspG1na1a	819
Db	4585	GGAGAAAGTTCCCTCCAGAACCAACCGGTTACTCTCCCTTGGCATGGCTTGTGATGAGA	4644
Qy	819	nVal1H1sPheAlaH1sThrSer1LeuLeuSerPheLeuVal1ArgG1aG1YrProTy	839
Db	4645	TGTTTCACTTGGACACCAAGATATATGCTGATTTTAAAGAGGCGCGGAAACCATTA	4704
Qy	839	rAspLeuG1n1LeYrProG1nG1naArg1a1aSer1LeaArgVal1ProG1uSerG1YrG1H1	859
Db	4705	TGACTTACAGATCTATCTCCAGAGAGACGACACGATCAGATTCCTGATCTCGAGAAC	4764
Qy	859	sTyG1uLeuH1sLeuLeuH1sTyTrpLeuG1nG1na1aSer1a1aG1YrArgG1Lea1a1a	879
Db	4765	CTATGAACCTGACCTGCTCACTCACTTACAGAGAACTTGGATCGCGTATCGTCTCT	4824
Qy	879	uLyVal1a1Le 882	
Db	4825	GAAAGTGATA 4834	
RESULT 3	AK050021	3327 bp mRNA linear HTC 05-DEC-2002	
LOCUS	AK050021	Mus musculus adult male liver tumor cDNA, RIKEN full-length	
DEFINITION	AK050021	enriched library, clone: C73003D12 product: DIPPTIDYL PEPTIDASE 9	
ACCESSION	AK050021	homolog [Homo sapiens], full insert sequence.	
VERSION	AK050021.1	GI:26340743	
KEYWORDS	HTC, CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
AUTHORS	Itch, H., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to		

JOURNAL	prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	20499374
REFERENCE	11042159
AUTHORS	3 Shibata, K., Itch, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
	Kono, H., Akizawa, J., Nishi, K., Kitanaka, T., Tashiro, H., Itch, M.,
	Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
	Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,
	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL	sequencing pipeline with 384 multicapillary sequencer
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	11076861
AUTHORS	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itch, M., Ishii, Y.,
	Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
	Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,
	Saito, T., Okazaki, Y., Gojodori, T., Bono, H., Kasukawa, T., Saito, R.,
	Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T.,
	Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
	Kuehl, P., Lewis, S., Matsuo, Y., Nakai, I., Pesole, G.,
	Quackenbush, J., Schriml, L.M., Staib, F., Suzuki, R., Tomita, M.,
	Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
	Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
	Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
	Hofmann, M., Hume, D.A., Kamuya, M., Lee, N.H., Lyons, P.,
	Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P.,
	Ring, B., Rindwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
	Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
	Toyo-Oka, K., Wang, K.H., Weitz, C., Whiteaker, C., Wilming, L.,
	Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S.
	and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	Nature 409 (6821), 685-690 (2001)
PUBMED	21085660
REFERENCE	11217851
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research
	Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL	of 60,770 full-length cDNAs
MEDLINE	Nature 420, 563-573 (2002)
PUBMED	6 (baes 1 to 3127)
REFERENCE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
AUTHORS	Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
	Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
	Hori, F., Imotani, K., Ishii, Y., Itch, M., Kagawa, I., Kasukawa, T.,
	Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
	Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
	Nakamura, M., Nishi, K., Nomura, K., Nomura, K., Ohno, M., Ohsato, N.,
	Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
	Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akita, S.,
	Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
	Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submersion
JOURNAL	Submitted (16-JUL-2001) Yoshinori Hayashizaki, The Institute of
	Physical and Chemical Research (RIKEN), Laboratory for Genome
	Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
	RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
	Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
	URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
	Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome
	Encyclopedia Project of Genome Exploration Research Group in Riken
	Genomic Sciences Center and Genome Science Laboratory in RIKEN.
	Division of Experimental Animal Research in Riken contributed to
	prepare mouse tissues.

Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.

Please visit our web site for further details.
 URL: <http://genome.gsc.riken.go.jp/>
 URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

SOURCE

1. 3327
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:C730003D12"
 /db_xref="FANTOM:10090"
 /clone="C730003D12"
 /sex="male"
 /cell_type="tumor"
 /tissue_type="liver"
 /clone_1ib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 176..2764
 /note="unnamed protein product; DIPPTIDY, PEPTIDASE 9 homolog (Homo sapiens) (SEPR|AAU47179, evidence: FASTY, 92%ID, 99.6%length, match=2580)
 putative"
 /codon_start=1
 /protein_id="BAC34034.1"
 /db_xref="GI:26340744"
 /translation="MCSGVSPVEOVAAGMDTAAFCVQKHSMDGLNSIHGRKSS
 GLIVSKAPDFOVQKPVDSGPHSHRLYLIMPYGSRENSLLYSSEIPKVRNALLL
 SKOMLDHFQATPHGVDSREBELLEKRLGVGITSDFSEGLFLFQSNLFLH
 CDGKNGFMVPMKPLIEIKTQCSGPMKPCPADPAFSPINSDLVANIERGE
 RLTFCHQSGAVLNPBKAGVATFVIOEFPKPCPDPAFSPINSDLVANIERGE
 EVDSEFVHTVPSPALERKTSRYRRTGSKPKALKAEIOTHQGTVSCE
 KEIUPFSSLPFKVEYIARAGTRDCKYAMAFIDRPOQLVLPLAFIPAYESE
 AQRQAARAVKNOVPFVIEEVTNWIMVHDIHPPQAEQDQFCFLRANEKCTGF
 CHLYRVTELKTKDYMTEPLSPTEDEPCPIKEEVALTSGEMVLSRGSXINWEO
 TLVYFQGTQKDPLEHNLVYVESAGEIVRLTLTGFSHSCSDFMFAHYSVS
 TPCVHYKLSGDDPDLKORFPMASMEANCPDVPYPEIPIHPTRADVOLGLM
 YKPHILOGRKHPVLFVYGGPOVOLVNNSPFGIKYLRILNTLASGYAVVLDGSC
 ORGLHFEALIKQKQKQVEIEDYEGIQYIAEKYFTIDSRVALHKSITGGFLSLGLI
 HKPQFKVALIAGAPVTVMVAIDTGTERTMDVPENNQCEKSGVALHVEKLPENPR
 LLIHGFLDENVHPHTNPLVSQLRAGKPYOLQIYPERHSIRCESEHVEYVTLH
 FLOEHL"

CDS

BASE COUNT 702 a 1015 c 919 g 691 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.45e-304 Length: 3327
 Score: 2833.00 Matches: 511
 Percent Similarity: 76.79% Conservative: 134
 Best Local Similarity: 60.83% Mismatches: 193
 Query Match: 60.28% Indels: 2
 DB: 11 Gaps: 2

US-10-070-464-1 (1-882) x AK050021 (1-3327)

QY 35 PheTyrValIGluArgTyrSerThrSerGlnLeuLysLysLeuAlaAspThrArgLys 54
 Db 242 TTCTGTGTGCAAGACACTCGTGGAGTGCGCGAGCATTAATCACGCGAGTGGCAAG 301
 QY 55 TyrHisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74
 Db 302 TCCCTGGGCGCTATTGTGTAGCAAGGCCGCCAGACTTCAGTTGTGTGAGAGGCTGAC 361
 QY 75 ProAspGlyProHisSerAspArgLysTyrTyrLeuAlaMetSerGlyGluAsnArgGlu 94
 Db 362 GAGTCTGGCCCCACCTTCACCGCTCTATTAATCTCGAATCCTTACCGCAGCGCTGAG 421
 QY 95 AsnThrLeuPheTyrSerGlnLysProLysThrIleAsnArgAlaAlaValLeuMetLeu 114
 Db 422 AACTCTCTCTACTCGAGATCCCAAGAAAGTGCGGAGAGAGCGCTGCTGCTGCTG 481

QY 115 SerTrpLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
 Db 482 TTCTGGAAACAGATGCTGGACCATTCACGGCCACACCCACCATGCTGCTACTCCGA 541
 QY 135 GluGluGluLeuLeuArgLysLysArgLysIleGlyThrValGlyIleAsnTyrAsp 154
 Db 542 GAGGAGGAGACTCTCGGAGGAGGCAAGCGCTGGCGGCTTCTGGAATCACTTATGAC 601
 QY 155 TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174
 Db 602 TTCCACAGTGAAGCGGCGCTTCTCTTCACAGGCCAGCAATAGCTGTTCACGTCAG 661
 QY 175 AspGlyGlyProGlnGlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSer 194
 Db 682 GATGCTGGCAAGATGCTTATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721
 QY 195 CysProAsnLysArgMetAspProLysLysCysProAlaAspProAspTrpIleAlaPhe 214
 Db 722 TCTTCTGGGCGCAGCATGGACCCCAAAATCTGCGGAGACCTGCTGCTTCTTCTTCT 781
 QY 215 IleHisSerAsnAspIleTrpIleSerAsnIleValThrArgGluGluArgLysLeuThr 234
 Db 782 ATCAACAAACAGATCTGTGGGTGGCAACATCGAGACTGGGAGAGAGAGCGGCTCAC 841
 QY 235 TyrValHisAsnGluLeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThr 254
 Db 842 TTCTGTACACAGGCTTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
 QY 255 PheValLysGlnGluGluPheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGlu 274
 Db 902 TTGTGATCATCAGAGAGAGTTCACAGGCTTCACTGGGTGCTGCTGCTGCTGCTGCT 961
 QY 275 ThrThrProSerGlyGlyLysLysLeuAlaGlyIleLeuTyrGluGluAsnAspGlySer 293
 Db 962 TGGAGAGCTCGAAGGCTCAAGACGCTGCAATCTATAGAGAACTGACGAGTCT 1021
 QY 294 GluValIGluIleIleHisValThrSerProMetLeuGluThrArgArgAlaAspSerPhe 313
 Db 1022 GAAGTGGAGTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
 QY 314 ArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGlyIleMet 333
 Db 1082 CGCTACCCACGAGACGAGCAGCAAGAACCCCAAGATGGCTGAGCTGAGCTGCTGAG 1141
 QY 334 IleAspAlaGluGlyArgIleIleAspValIleAspLysGluLysIleGlnProPheGlu 353
 Db 1142 ACGACCATCAGGCAAAATCTGTCAAGCTCGAAGAGAACTGTACAGCATTCACAC 1201
 QY 354 IleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTyrTrpThrProGluGlyTyr 373
 Db 1202 TCCCTTTTCCCAAGTGAAGTGAATGCGCCGAGCTGCTGAGACACGAGGAGCAATAT 1261
 QY 394 AlaTrpSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393
 Db 1262 GCTGCGGCATGTTCTGTGAGCGCTGCCAGCAAGCGCTTACGCTTCTGCTGCTGCT 1321
 QY 394 GluLeuPheIleProValIGluAspAspValMetGluArgGlnArgLeuIleGluSerVal 413
 Db 1322 GCTCTTCTATCTCCGCGGTGAGAGTGAAGGCCCGCGGACGAGCTGCCAGACCGCT 1381
 QY 414 ProAspSerValThrProLeuIleIleTyrGluGluGluThrAspIleTrpIleAsnIle 433
 Db 1382 CCCAAGATGTGACAGCCCTTGTCTATGTAGAAAGTCAACCAATGTCTGATCAACGTC 1441
 QY 434 HisAspIlePheHisValPheProGlnSerHisLysGluGluGluIleGluPheIlePhe 452
 Db 1442 CACGACATCTTCAACCGCTTCTCAGGCTGAGGCGCAGAGACTTTTGTGTTCTTGTGT 1501
 QY 453 AlaSerGluCysLysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGlu 472
 Db 1502 GCCAAGATGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1561
 QY 473 SerLysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIle 492

```

Db      1562 AAGAGCTAGTACTGAGGAAACCCCTCAGCCCTTACAGAAAGTAGTTTAAGTCCCATC 1621
Oy      493  LysGluGluIleAlaIleThrSerGlyGluTyrGluValIleuGlyTyrGhiAglySerAa 512
Db      1622 AAGAGAGAGGTGCCCTGACCAATGCGAGTGGAGAGCTCTTCGAGGCAATGGCTCCAA 1681
Oy      513  IleGluValAspGluValArgArgLeuValTyrPheGluGluTyrThrLysAspSerProLeu 532
Db      1682 ATCTGGGTCAACGAGAGAGCAAGCTGGTGTACTTCAAGAGTCAAAAGGACACACCGCTG 1741
Oy      533  GluHisIleLeuTyrValValSerTyrValAsnProGluGluValIleArgLeuThrAsp 552
Db      1742 GAACATCACCCTCTATGTGTGCTGACACTGAGTACAGGCGAGATCTCGCGCTCACCACAG 1801
Oy      553  ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPheIleSerTyrTyr 572
Db      1802 CTCGGCTTCCCAAGAGCTGCTCCATGAGCCACAGCTGACATGTTCGGAGTCACTAC 1861
Oy      573  SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 592
Db      1862 AGCAGATGAGCAGCGCCACCTGTGTACATGTGTACAGCTGAGCGCGCCGATGATGAC 1921
Oy      593  ProThrCysLysThrLysGluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuPro 612
Db      1922 CCATCTGACAGAGCAACGCTTCTGGCCGACATGTGAGGAGGAGCAATTCGCCCA 1981
Oy      613  AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632
Db      1982 GACTATGTGCCCTGAGATCTTCCACTTCCACACCGGTGCACACGTGACGCTCTACGG 2041
Oy      633  MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652
Db      2042 ATGATCTACAGAGCACACACCCCTGCAACCTGGAGAGAGCAACCCCTGTGCTTGTGTC 2101
Oy      653  TyrGlyGlyProGluValGluLeuValAsnAspArgPheIleGlyValLysTyrPheArg 672
Db      2102 TATGGGGGCCACAGGTGCTTGGTGGTGAACACTCTTTAAGGGCATCAATACCTCGCG 2161
Oy      673  LeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspAsnArgGlySerCys 692
Db      2162 CTAAATACACTGGCATCTTGGGCTATGCTGTGTGTATCATGTGTGCGGGCTCTCTG 2221
Oy      693  HisArgGlyLeuLysPheGluGluValAlaPheLysTyrLysMetGlyGlnIleGluLeuAsp 712
Db      2222 CAGCGGGGCTGCACCTTCCAGGGGGGCTTGAATAATCAATGGGCCAGGGGAGATTGAG 2281
Oy      713  AspGluValGluGluLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732
Db      2282 GACCAAGTGGAGGCTTGCAGTACGTGGCTGGAAGATATGGCTTCACTTGAACCCCA 2341
Oy      733  ValGlyIleHisGlyTyrPheTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
Db      2342 GTCCGATCATGCTGCTGCTCTCCGCGGCTTCTCTCACTCATGGGGCTCATCCACAAG 2401
Oy      753  SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAsp 772
Db      2402 CCAAGAGTGTCAAGGTAGCCATATGCGGGGCTCTGTCACTGTGTGAGAGGGCTTATGAC 2461
Oy      773  ThrGlyTyrThrGluArgTyrMetGlyHisPheProAspGlnAsnGluGlnGlyTyrTyrLeu 792
Db      2462 ACGAGGTACACGGAAGATACATGATGTCCCGAAATAACACAGCAAGGCTTATGAGCA 2521
Oy      793  GlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeu 812
Db      2522 GGATCTGTAGCCCTCATGTGAGAGAGCTGCCAATAGAGCTTAAACCGCGCTTATCTTC 2581
Oy      813  HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
Db      2582 CACGGGCTTCTGAGCAAGAGCTTCACTTCTTCCACAAATTTCTTGGGTCTCCAGCTG 2641
Oy      833  ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852

```

```

Db      2642 ATCCGAGCAGGAAGCCATATCAAGCTTACATCTTCAACCAAGACAGATACATCCG 2701
Oy      853  ValProGluSerGlyGluHisTyrGluLeuHisIleLeuHisTyrLeuGlnGluAsnLeu 872
Db      2702 TCCCGGAGATCCCGAGAGCATTAACAGATGACGCTCTCTCACTTCTGACAGAACACCTG 2761

RESULT 4
AK078301
LOCUS
DEFINITION
Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430584g1 product:DIPEPTIDYL PEPTIDASE 9
homolog [Homo sapiens], full insert sequence.
ACCESSION
AK078301
VERSION
AK078301.1 GI:26347124
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
JOURNAL
MEDLINE
PUBMED

TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20493374
11042159
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Komoto, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
REFERENCE
AUTHORS
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Komoto, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staudil, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Balarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, N., Gustincich, S., Hill, D.,
Hofman, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombereite, P., Nordone, P.,
Ring, B., Rongwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

```

TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL	of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002)
AUTHORS	6 (bases 1 to 3457)
	Adachi, J., Aizawa, K., Akimura, T., Bono, H., Carninci, P.

TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan [E-mail: genome-res@gscc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>.

CDS

FEATURES

source

BASE COUNT	766 a	1033 c	926 g	732 t
ORIGIN				
Alignment Scores:				
pred. No.:	2.04e-303		length:	3457

Score:	2823.00	Matches:	510
Percent Similarity:	76.67%	Conservative:	134
Best Local Similarity:	60.71%	Mismatch:	194
Query Match%	60.06%	Indels:	2
DB:	11	Gaps:	2
US-10-070-464-1 (1-882) x AK078301 (1-3457)			
QY	35 PheTYrValGluArgTYrSerTRPserGlnLeuLYsLYsLeuAlaAspThrArgLYs	54	
DB	248 TTCTGTGTGCAAGAAAGCACTGTGGATGGGGCTGGCGATGATTATCCAGCGAGCTGCAAG	30	
QY	55 TYrHisGLYrTYrMeTcAlaLYsAlaProHisAspPheMeTcPheValLYsArgAsnAsp	74	
DB	308 TCCCTGGGCGCTCATTTGTGACGAAGGGCCCCCGACGACTTCAGTTTGTGCAGAACGCTGAC	36	
QY	75 ProAspGLYProHisSerAspArgLYrTYrTYrLeuAlaMeTserGLYAsnArgLYu	94	
DB	368 GAGTCTGGCCCCCACTCTACCGCTCTATTAACCTCGGAATGCTTACCGGACCGCTGAG	42	
QY	95 AsnThrLeuPheTYrSerGlnLYrProLYsThrLYsAsnArgAlaLYsAlaLeuMeTcLeu	11	
DB	428 AACCTCCCTCTCTACTTCGAGATCCCAAGAAAGTCGGAAGAGAGCCCTGCTGCTG	48	
QY	115 SerTRYPLeProLeuLeuAspLeuPheGlnAlaThrLeuAspTYrGLYMeTYrSerArg	13	
DB	488 TCCCTGAAGACAGATGCTGACCACTTCAGGCGACACCCACCATGAGTGTACTACCCGA	54	
QY	135 GlnGlnGlnLeuLeuArgLYuArgLYuAsaGLYrHisTYrValGLYrLYsAlaSerTYrAsp	15	
DB	548 GAGAGAGAGCTACTGCGGAGGCGCAAGCCCGGCGCTTCCTGGAATACCTCTTATGAC	60	
QY	155 TYrHisGLNLYsSerGLYrPheLeuPheGlnAlaGLYserGLYrLYrHisValLYs	17	
DB	608 TTCCACAGTAGAGAGCGGCTCTTCTCTTCCAGGCGAGCAATAGCTGTTCACCTGCAAG	66	
QY	175 AspGLYGLYProGlnLYrPheThrGlnLYrProLeuArgProAsnLeuValGlnThySer	19	
DB	668 GATGTGTGCAAGAAATGGCTTTATGAGTGTCCCGATGAAGCCACTGAGATCAAGACTGAG	72	
QY	195 CysProAsnLYrLeuArgMetAspProLYsLeuCYsProAlaAspProAspTRPLeuAlaPhe	21	
DB	728 TGTCTGTGGGCGACGAGATGACCCCAAAATCTGCCCGAGACCCCTGCTCTTTTCCCTTC	78	
QY	215 ILeHisSerAsnAspLYrLeuTrpLYsSerAsnLYsValThrArgGlnGlnArgLYuLeuThr	23	
DB	788 ATCAACAACAGTATATCTGTGGGTGGCAACATCGAGACTGGGGAGAGAAAGCGGCTCAC	84	
QY	235 TYrValHisAsnGlnLeuAlaAsnMetGlnGlnAlaAspAlaArgSerAlaGLYValAlaThr	25	
DB	848 TTCTGTACCAAGGGTTCACTGGTGTCTCTGACCAATCCCAATACGACGAGGCTGGCAC	90	
QY	255 PheValLeuGlnGlnGlnPheAspArgTYrTYrTRPProLYsAspLYsAlaGlu	27	
DB	908 TTGTGATCCAGAGAGAGATTGACCGCTTCACTGGGTGCTGGTGGTGCACCGGCTCT	96	
QY	275 ThrThrProSerGLYLYsLYsLYsLeuArgLYrLYsGlnGlnGlnAsnAspGlnSer	29	
DB	968 TGGGAAGGCTCCCAAGGCTTCAAGACGCTGCGCATCTCATATGAGAAAGTGACGAGCT	102	
QY	294 GlnValGlnLYrLeuHisValThrSerProMeTcLeuGlnTrpArgArgAlaAspSerPhe	31	
DB	1028 GAAGTGAAGTCAATTCATGTGCGCTTCCCGCGCTCGAGAGAGAGAAACGACTCTTAC	108	
QY	314 ArgTYrProLYrThrGLYrThrAlaAsnProLYsValThrPheLYsMeTserGlnLeuMet	33	
DB	1088 CGCTACCCCAAGGACAGGACGAAGAAACCCCAAGATTGGCTCGAAGCTGCTGACCTCAG	114	
QY	334 ILeAspAlaGlnGlnLYrArgLYrLeuAspValLYsAspLYsGlnLeuLYrGlnProPheGln	35	
DB	1148 ACGGACCATCATGGGCAAAATCGTGTCAAGCTGCGGAAGAACTGTGACCACTTACG	120	
QY	354 ILeuLeuPheGlnGlnLYrValGlnLYrLYrLeuAlaArgAlaGLYrTRPTrpProGlnGlnLYsTYr	37	

Db 1208 TCCCTTTTCCCAAGGAGTACATGCGCCGCGTGGACAGCGGCAATAT 1267
 Qy 374 AATPSeTLeuLeuAAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPro 393
 Db 1268 GCTGGGCGCATGTCCTGTATCCGTCGCCAGCAAGGCTTAGCTTCTCCCAACCCCT 1327
 Qy 394 GluLeuPheIleProValGluAAspArgValMetGluArgGlnArgLeuIleGluSerVal 413
 Db 1328 GCTCTCTTATCCCGCGTGGAGTGAAGGCGCAGCGGCGACGCTGCCAGACGCGTC 1387
 Qy 414 ProAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIle 433
 Db 1388 CCCAAGAAATGAGCCCTTGTATCATCTATGAAAGAGTCAACATGCTGATGATCAAGTC 1447
 Qy 434 HisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePhe 452
 Db 1448 CACGACATCTTCCACCCGTTTCTCAAGCTGAGGCGCAGACAGACTTTGTTCTCTTCG 1507
 Qy 453 AlaSerGluCysAlaSerThrGlyPheArgHisLeuTyrIleTyrSerIleLeuLeuArg 472
 Db 1508 GCCAAGAAATGAAAGATGGCTTCTGCCACTGTATACAGGCTCACAGTGAACTTAAAC 1567
 Qy 473 SerLeuTyrLeuArgSerSerGlyGlyLeuProAlaProSerAspPheLeuSerProIle 492
 Db 1568 AAGGACTATGACTGAGACGGAACCCCTCAGCCCTACAGAAAGATGATTTAAGTCCCATC 1627
 Qy 493 LysGluGluIleAlaIleThrSerGlyGluTyrGluValLeuGluValArgHisGlySerAsn 512
 Db 1628 AAGGAGAGAGTGCGCTGACACAGTGGAGTGGAGAGTCTTTCGAGGCGATGCTCCAG 1687
 Qy 513 IleGlnValAspGluValArgArgLeuValTyrPheGluGlyTyrThrAspSerProLeu 532
 Db 1688 ATCTGGGTAAACAGAGCAAGCAAGAGCTGGGTGTACTTTCAAGGTACAAAGACACACCGCG 1747
 Qy 533 GluHisHisLeuTyrValIleSerTyrValAsnProGlyGluValThrArgLeuThrAsp 552
 Db 1748 GAAACATCACTCTATGTGGTCAAGCTACAGACTCAGACAGGCGAGATCTGGCGCTCACCA 1807
 Qy 553 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyr 572
 Db 1808 CTCGGCTTCTCCACAGCTGCTCCATGAGCCAGAGCTTGACATGTGTTGAGTCACTAC 1867
 Qy 573 SerAsnGlnLeuAspProHisCysValSerLeuTyrLeuLeuSerSerProGluAspArg 592
 Db 1868 AAGAGGTGAGACACGCGACCTGTGTACATGTGTACAGCTGAGCGGCCCATGTATGAC 1927
 Qy 593 ProThrCysLeuThrLysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuPro 612
 Db 1928 CCACGTGCAAGAGAACACAGCTTCTGGCGCAGCATGATGAGAGCGACCAATTGCCCCCA 1987
 Qy 613 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 632
 Db 1988 GACTATGTGCCCCCTGAGATCTTCCACTTCCACACCCGTCAGACCTGACGCTTACGCG 2047
 Qy 633 MetLeuTyrLeuProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 652
 Db 2048 ATGATCTTACAGCCACACACCTGCAACTTGGAGAGAACCCCACTGTGCTCTTGTTC 2107
 Qy 653 TyrGlyGlyProGlnValGlnLeuValAsnAspArgPheLysGlyValLysTyrPheArg 672
 Db 2108 TATGGGGGCCACAGAGTGCAGTTGTGGAACAATCTCTTAAGGGCATCAAAATACCTGGG 2167
 Qy 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCys 692
 Db 2168 CTAAATATACCTGACCTTGGGCTATGTCTGTGTGTATGATGATGTGTGGGGCTCTCTGT 2227
 Qy 693 HisArgGlyLeuLeuPheGluGlyAlaPheLysTyrLeuMetGlyGlnIleGluLeuAsp 712
 Db 2228 CAGCGGGGCTGCACTTCAAGGGGGCGCTGAAAAATCAAAATGAGCGAGATGAG 2287
 Qy 713 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732

Db 2288 GACCAAGTGAAGAGCTTGCAATACGTGCTGAGAAAGATATGGCTTACTTACCTTGACCCCA 2347
 Qy 733 ValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
 Db 2348 GTTCGCATCATAGGCTGGTCTTACAGCGGCGCTTCTCTCATCTATGAGGGCTCATCCAAAG 2407
 Qy 753 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAsp 772
 Db 2408 CCACAAGTTCACAGGTACCAATGGCGGCGCTCTGTCATCTGTGTGATGCGCTATGAC 2467
 Qy 773 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGluAsnGluGlnGlyTyrTyrLeu 792
 Db 2468 AAGAGGTACACCGAAGCAATACATGATGATCCCGAAATATACAGAAAGCTATAGGCA 2527
 Qy 793 GlySerValAlaMetGlnAlaGluArgPheProSerGluProAsnArgLeuLeuLeu 812
 Db 2528 GGATCTGTAGCCCTGCATGTGAGAAAGTGGCCCAATGAGCTTAACCGCTGCTTATTC 2587
 Qy 810 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
 Db 2588 CACGGCTTCTGAGAGAGAACGTTCACTTCCACACAAATTTCTGTGTGCTCCAGCTG 2647
 Qy 833 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 852
 Db 2648 ATCCGAGCGAAGAACCATACCACTTCAATCTACCCAAAGAGAGATAGATCCCG 2707
 Qy 853 ValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 872
 Db 2708 TCCCGGAGATCCGAGAGAGATTACAGAGGTGAGACGCTGCTCACTTCTGACAGAAACACTG 2767
 RESULT 5
 AK050023
 LOCUS
 DEFINITION
 AK050023 3376 bp mRNA linear HTC 05-DEC-2002
 Mus musculus adult male liver tumor cDNA, RIKEN full-length
 enriched library, clone: C73003H19 product: D1PEPTIDYL PEPTIDASE 9
 homolog (Homo sapiens), full insert sequence.
 AK050023
 VERSION
 AK050023.1 GI:26093756
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS
 Carninci, P. and Hayashizaki, Y.
 TITLE
 High-efficiency full-length cDNA cloning
 JOURNAL
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE
 99279253
 PUBMED
 10349636
 REFERENCE
 AUTHORS
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE
 20493374
 PUBMED
 11042159
 REFERENCE
 AUTHORS
 3 Shibata, K., Itoh, M., Atzawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M.,
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed capillary sequencer
 JOURNAL
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE
 20530913
 PUBMED
 11076861
 REFERENCE
 AUTHORS
 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,

QY	313	EAISGTyrProHisThrArgIleValAlaAspProValLeuThrPheMetSerGluIle	333
Db	1141	CCGGTACCACAGGACAGCAGCAGAACACCCAGATTTCCTGTAAAGCTGGCTGACCTCA	1200
QY	333	CTLEAspAlaGluArgIleIleAspValIleAspValGluLeuIleGlnProPheG1	353
Db	1201	GACGACCATCAGGGCAAAATTCGTGTCAAGCTGGAGAAAGAACTGTGACGCCATTCAG	1260
QY	353	uILEuPheGluGluValIleGluTyrIleAlaArgAlaGluTyrPheProGluGlyValTyr	373
Db	1261	CTCCCTTTTCCCAAGAGGAGATCATGCCCGGGCTGGCTGGACACCGGAGGCCAATA	1320
QY	373	AlATrSerSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerPr	393
Db	1321	TGCCCTGGGCCATGTTCTCTGAGACCGTCCCAAGACGGCTTCAAGCTTCTCTGCCCC	1380
QY	393	oGluLeuPheIleProValGluAspAspValMetGluArgGluArgLeuIleGluSerVa	413
Db	1381	TGCTCTCTTCAATCCGGCGGTGTGAGAGAGAGGCCCCAGGCGGACGCTGCCAGAGCCGT	1440
QY	413	lProAspSerAlaThrProLeuIleIleTyrGluGluThrThrAspIleTyrPleAsnI1	433
Db	1441	CCCCAAGATGTGCAGCCCTTGTGTATCTTAAGAAAGTCAACCAATGCTGGATCAACGT	1500
QY	433	eHisAspIlePheHisValPheProGlnSerHis---GluGluGluIleGluPheIlePh	452
Db	1501	CCAGACATCTTCCACCCGTTTCTCCAGAGCTGAGGGCCAGCAGAGACTTGTGTTCTTCG	1560
QY	452	eAlAspSerGluCysLeuThrArgIlePheArgHisLeuTyrIleValIleThrSerIleLeuValG1	472
Db	1561	TGCCCAACAAATGCAGAGACTGGCTTCTTGCACCTGTACAGGGCTCAACAGAACTTAAAC	1620
QY	472	uSerIleTyrIleAspSerSerGlyGlyLeuProAlaProSerAspPheLysCysProI1	492
Db	1621	CAAGGACATGATGATGACGAGCAACCCCTCAGCCCTACAAAGATAGATTAACTGCCAT	1680
QY	492	eLysGluGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAs	512
Db	1681	CAAGAGAGAGGTCCGCTCGACACAGTGGGAGGTGGAGGTCTTGTTCGAGCATGTGCTCCA	1740
QY	512	nIleGlnAlaAspGluValArgArgLeuValTyrPheGluGlyHisPheAspSerProLe	532
Db	1741	GATCGGGTCAACGAGCAGACGACGACTGTGATCTTAAGGTAAAGGACACACCGCT	1800
QY	532	uGlnHisIleLeuTyrValIleSerTyrValAsnProGlyGluValThrArgLeuThrAs	552
Db	1801	GGAACATCACTCTATGTGTGTGATGACGTAAGAGCAGACAGCCAGATCGTCCGGCTCACAC	1860
QY	552	pArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerValTyr	572
Db	1861	GCTGGGCTTCTCCACACACTCTCTCATAGCCAGGACTTCGACATGTTGCGAGTCACTA	1920
QY	572	rSerAsnGlnLysAsnProHisCysValSerLeuTyrIleLeuSerSerProGluAspAs	592
Db	1921	CAGCAGTGTAGACACGCCACTGTGTACATGTGTACAACTGACGAGCCGCCCGATATGA	1980
QY	592	pProThrCysIleThrLysGluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuPr	612
Db	1981	CCCACTGCACAAAGCAACACCGCTTCTGGGGCCAGCATGTATGAGGACGACCAATTGCCCCC	2040
QY	612	oAspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrG1	632
Db	2041	AGACTATGTGCCCTCGAGATCTTTCACCTTCCACCCGTGCAGACGTCGACGCTCTACGG	2100
QY	632	uMetLeuTyrLysProHisAspLeuGlnProGlyLysIleTyrPProThrValLeuPheI1	652
Db	2101	CATGATCTACAAGCACACACCTCTCAACTGTGGAGGAAGACCCCACTGTGCTCTTGT	2160
QY	652	eTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheAr	672
Db	2161	CTATGGGGGCCCAACAGGTGATGGTGTAAACAATCTTTAAGGGCAACAATATACCTGGC	2220

[illegible]

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5466759"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC library."
BASE COUNT      279 a      223 c      214 g      303 t      3 others
ORIGIN
Alignment Scores:
Pred. No.:      5,23e-178      Length:      1042
Score:          1693.50      Matches:      324
Percent Similarity: 93.97%      Conservative: 3
Best Local Similarity: 93.10%      Mismatches: 15
Query Match:    36.03%      Indels:      6
DB:              12      Gaps:      1
US-10-070-464-1 (1-882) x BM557438 (1-1042)
OY 504 GUAUALLUAGUAYAGHISGYSERANILLEGINVALASPGUVALARGHLEUVALTYR 523
DB 1 GAAGTTCTTGCGCGGCATGATCTAATATCCAAAGTGAAGTACGAAGGCTGGATATAT 60
OY 524 PHEGUGUJYTHRYLADSPSERPROLEUGIHSHILEUTYRVAVASERYTYRVALAN 543
DB 61 TTGAAGGACACCAAGACTCCCTTTAGACATCACTGTAGTACGATCACTTAATAT 120
OY 544 PROGLYUVALTHARGLEUTHASPARAGLYTYRSEHISSECYESEYILLESEGIN 563
DB 121 CCTGAGAGGTACAAAGCTGACGACCGTGGCTACTCATCTTTGCGCATCACTGAG 180
OY 564 HISCYASAPHEPHEILSEERYTYRSEANGILNLYASAPROHICCYAVALSERLEU 583
DB 181 CACTGTGACTCTTTATAGTAGTAGTAACCAAGAAATCCACACTGTGTCCCTT 240
OY 584 TYRVLASERSESPROGLUASPARPROTHRYSLYSTRYGLUHPHETPAATHR 603
DB 241 TACAAGCTATCAAGCTCGAAGATGACCAACTTGCAAAAAGAAATTTGGGCCACC 300
OY 604 ILEUASPSERAJAGLYPROLEUPROASPTYRTHRPROPROGLUULEPHESERPHGLU 623
DB 301 ATTTTGGATTACAGCAGGTCTCTTCTGACTATCTCCCTCCGAAATTTTCTTTTGA 360
OY 624 SERTHRTHRYLPHETHLEUTYRGLYMECTLEUTYRYSPROHISASPLEUGINPROGLY 643
DB 361 AGTACTACTGATTTACATTGATGAGATGCTCTCAAGCCCTCATGATCTACAGCTGGA 420
OY 644 LYSLSYSTRPROTHVALLEUPHEILETYRGLYGLYPROGLINVALGLEUVALASPHEN 663
DB 421 AAGAAATATCCACTGTCTGTCTATATATGAGGCTCCAGGTGAGTGTGATATAT 480
OY 664 ARGPHELYSGLYVALLYSTRYPHEARGLEUASPHNTHLEUALASERLEIGLYTYRVALVAL 683
DB 481 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCTTAGCCTCTAGGTTATGTGTT 540
OY 684 VALVALILEASPARHARGLYSERCYSHISATAGLYLEULYRPHGLUGLYVALAPHELYS 703
DB 541 GTAGAGATAGCAACAAGGGGATCCCTGTCAACCAAGGGCTTAATTTGAAGGGCCCTTTTMA 600
OY 704 TYRYSMETGLYGINILEGLUULEASPARGLINVALIGLYLEUGINLYTYRLEUALASER 723
DB 601 TATTAATATGGGTCAATATGAATTTACATCAAGGTGAGAGGACTCAATATCTTACTCT 660
OY 724 ARGTYRASPHLEILEASPLEUASPARVALYGLYLEHISGLYTRIPSERYTYRGLYTYR 743

```

```

DB 661 CGATATGATTTTCATTGACTTATGCTGTGGGCATCCACGGCTGGTCCATGAGAGATAC 720
OY 744 LEUSERLEUMETALILEUMETGLNARGSERAPILIEPHARGVALAIALEAAGLYALA 763
DB 721 CTCTCCCGATGCGCATTAATGACAGAGTCAGATATCTTCAGGGTTGCTATTCCTGGGGCC 780
OY 764 PROVALTHRYLLEUTRILPHETHRAPHHRGRYTYRTHGLUATYRTHMETGLYHISPRO 783
DB 781 CCAGTCACTCTGTGATCTCTTATATACAGATACACGGTACGTTATATATGAGTACCTT 840
OY 784 ASPGINANGILUGLNGLYTYRTRYLEUGLYSERVALAIAAMEGLNALAGLUYSPHEPRO 803
DB 841 GACCGAATAGAACAGGGCATATTACTTANGATCTGTGGCATATCAAGCAAGAAAGTCCCCC 900
OY 804 SERGIUPROASPARHARGLEULEULEUHIISGLY-PHELEUASPGUASPHVAL-HISPHEA 823
DB 901 TCTGAACCAATCGATTAATCTGCTTTACATGCTTTTCTCGATGAGAAATGTCACATTTTG 960
OY 823 LAHISHSER-ILEULEUSERPHELEUVALARGALAGLYLYPEPROTHRYRAPHLEU--- 841
DB 981 CACATACCCAGAGATATTAAGTATTTTNTTGAAGGGGCTGAAAAAGCCATATGAT 1020
OY 842 -----GlnIleTYRPro 845
DB 1021 TTACACCAATCTATCTT 1038
RESULT 7
BO675006 957 bp mRNA linear EST 15-JUL-2002
LOCUS BO675006
DEFINITION AGENCOURT_8303592 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6274622
5' mRNA sequence.
VERSION BO675006
KEYWORDS BO675006.1 GI:21785840
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 957)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM2456 row: h column: 15
High quality sequence start: 4
High quality sequence stop: 624.
FEATURES
location/Qualifiers
1..957
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6274622"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 102"
/note="Organ: salivary gland; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Library constructed
by Ling Hong in the Laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC library."
BASE COUNT      257 a      206 c      228 g      266 t

```

ORIGIN

Alignment Scores:

Pred. No.: 3,95e-161 Length: 957
 Score: 1542.00 Matches: 298
 Percent Similarity: 95.24% Conservative: 2
 Best Local Similarity: 94.60% Mismatches: 10
 Query Match: 32.81% Indels: 5
 Gaps: 0

US-10-070-464-1 (1-882) x BQ675006 (1-957)

QY 474 LysTyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLys 493
 DB 12 AAATATAAAGATCCAGTCGTCGGCTGCTCCAGGATTTTCAAGTCTCTTATCAAA 71
 QY 494 GluGluAlaAlaIleThrSerGlyGlyTyrGluValLeuGlyArgHisGlySerAsnIle 513
 DB 72 GAGGAGATGACAAATTCAGTGGTGAATGGGAAGTCTTGGCCGCAATGATCTTAATATC 131
 QY 514 GluValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlu 533
 DB 132 CAAGTTGATGAAGTCAAGAGGCTGGTATATTGAAAGGACCAAGAGCTCCCTTTAGAG 191
 QY 534 HisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArg 553
 DB 192 CATCACCTGTACCTAGTACGTTACGTAAATCCGAGAGGCTGCAAGGCTGACCTGACCGT 251
 QY 554 GlyTyrSerHisSerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSer 573
 DB 252 GGCTACTACATCTTCTGCTGATCAGTCAGCTGAGCTGACTCTTTATTAAGTAAATAGT 311
 QY 574 AsnGluLysAsnProHisCysValSerLeuTyrLysLeuSerProGluAspAspPro 593
 DB 312 AACCAAGAAGATCCACACTGTGTCTCTTACAGATCAAGTCTCTGAAAGATGACCCA 371
 QY 594 ThrCysLysThrLysGluPheTyrPheIleLeuAspSerLysGlyProLeuProAsp 613
 DB 372 ACTTGGAAACAAAGAAATTTTGGCCACCATTTTGGATTCACAGGCTCTCTTCTGAC 431
 QY 614 TyrThrProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 633
 DB 432 TAACTCTCCAGAAATTTCTCTTTGAAAGTACATGATTTACATTTATGGAGATG 491
 QY 634 LeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyr 653
 DB 492 CTCTACAGGCTCATGATCTACAGCCTGGAAGAAATATCTCTGCTGCTTCATATAT 551
 QY 654 GlyGlyProGluValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 673
 DB 552 GGTGGTCTCCAGGTGACGTTGTAATATCGGTTTAAAGAGATCAAGTATTTCCGCTTG 611
 QY 674 AsnThr-LeuAlaSerLeuGlyTyrValValValAlaIleAsp-AsnArgLysSerCysH 693
 DB 612 AATACCCCTAGGCTCTCTAGTATATGTTGTTAGTATGACAAACAGGGAGATCTGTC 671
 QY 693 IsetrGlyLeuLysPheGlyValAla-PheLysTyrLysMet-GlyGlnIleGluIleAs 712
 DB 672 ACCGAGGGCTTAATATTTGAAGGCGCCTTTAATATTAATATGAGGCAATATAGAAATTTGA 731
 QY 712 PheArgIleValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAsp 732
 DB 732 CGATTCAGGGGAGAACTCCAAATATCTAGCTTCCGATATGATTTCAATGATTAATATCG 791
 QY 732 GValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
 DB 792 TGTGGGCAATCAAGCGGTGGCTTATGGAGATCCCTCTCCGTAAGGCAATTTATGCAAG 851
 QY 752 GSerAspIlePheArgValAlaIleAlaGlyAlaProValThr-LeuTyrIlePheTyr 772
 DB 852 GTCAAGATATCTTCAGAGTTCATTTGCTGGGGGCCAGTCACATCCGTTGATCTTCATG 911
 QY 772 spthrGlyTyrThrGluArgTyrMetGlyHisPro 783

DB

912 ATACAGATACCCGGAACGTTATATGAGGCCCCC 946

RESULT 8 BQ675260 910 bp mRNA linear EST 15-JUL-2002
 LOCUS BQ675260 AGENCOUNT 8354972 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275478
 DEFINITION 5', mRNA sequence.

ACCESSION BQ675260 GI:21786094
 VERSION BQ675260
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 910)
 NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Prepared by: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM2458 row: 1 column: 07
 High quality sequence stop: 618.

FEATURES

source

1..910
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6275478"
 /issue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH_MGC_102"
 /note="Organ: salivary gland; Vector: pOT87; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming;
 directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-CDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH_MGC library."

BASE COUNT

248 a 186 c 212 g 262 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 6.9e-161 Length: 910
 Score: 1539.50 Matches: 293
 Percent Similarity: 97.04% Conservative: 2
 Best Local Similarity: 96.38% Mismatches: 5
 Query Match: 32.76% Indels: 4
 Gaps: 1

US-10-070-464-1 (1-882) x BQ675260 (1-910)

QY 475 TyrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGlu 494
 DB 2 TATTAACGATTCAGTGGTGGCTGCTCCAGAGATTTCAAGTCTCTATCAAAAG 61
 QY 495 GluIleAlaIleThrSerGlyGlyTyrGluValLeuGlyArgHisGlySerAsnIleGln 514
 DB 62 GAGATGCAATTAACAGTGTGAATGGAGAGTCTTGGCCGCAATGATCTAATATCCA 121
 QY 515 ValAspGluValArgArgLeuValTyrPheGlyGlyTyrLysAspSerProLeuGluHis 534
 DB 122 GTTGATGAAGTCAAGAGCTGGTATATTTGAAAGCAACAAGAGATCCCTTTAGAGCAT 181
 QY 535 HisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGly 554

```

Db      |||||
182  CACCTGTAGTGTAGTACGTAAATCCTGGAGAGGTGACAAAGCTGACCTGCGTGC 241
Qy      |||||
555  TYSerHisSerCyCyAlIleSerGlnHisCyAaspPheHeIleSerIyYrSerAan 574
Db      |||||
242  TACTACATCTTCTGTCATCAGTCAGCAGCTGACCTCTTTATATAGTAAGTACTATAC 301
Qy      |||||
575  GlnLysAspProHisCyValSerLeuYrYrLysLeuSerSerProGluAspAspProTr 594
Db      |||||
302  CAGAAAGATCCACACTGTGTGTCCTTTACAGCTTCAAGTCTTAAGATGACCCAACT 361
Qy      |||||
595  CyLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspYr 614
Db      |||||
362  TCCAAACAAAGAAATTTGGGCCACATTTGGATTGACAGGTCTCTTCTGACTAT 421
Qy      |||||
615  ThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuYrGlyWetLeu 634
Db      |||||
422  ACTCTCCGAAATTTCTTTTGAAGTACTACGAGATTAACATTTGATGGAGCTTC 481
Qy      |||||
635  TyrLysProHisAspLeuGlnProGlyLysLysYrProThrValLeuPheIleYrGly 654
Db      |||||
482  TCAAAAGCTTCATGATTTACAGCTGGAAGAAATATCTTACTGTCTGTCATATATGAT 541
Qy      |||||
655  GlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysYrPheArgLeuAsn 674
Db      |||||
542  GGTCTTCAGGTGACAGTTGGTGAATATCGATTAAAGAGTCAAGTATTTCCGCTTGAAT 601
Qy      |||||
675  ThrLeuAlaSerLeuGlyYrValValValIleAspAsnArgGlySerCyHisArg 694
Db      |||||
602  ACCCTAGCTCTCTAGGTATGTTGTTAGTATGACAAACAGGGGATCTGTCTACCA 661
Qy      |||||
695  GlyLeuLysPheGlnGlyValAlaPheLysYrYrLysMetGlyGlnIleGluIleAspAspGln 714
Db      |||||
662  GGGCTTAATTTGAAAGGCCCTTTAAATATTAATATGATGATGAAATATGACATCAG 721
Qy      |||||
715  ValGlnGlyLeuGlnIleLeuAlaSerArgYrYrAspPheIleAspLeuAspArgValGly 734
Db      |||||
722  GTGAAGAGACTCCATATATCTAGCTTCTCGATATGATTTGATGACTTATGATGCTGGGC 781
Qy      |||||
735  IleHisGlyTrpSerYrGlyYrYrLysSerLeuMetAla-LeuMetGlnArgSerAan 754
Db      |||||
762  ATCCAGGCTGATCCATAGAGATACCTCTCCGATGATGATTTAAAGTCAAGAGCAAA 841
Qy      |||||
754  P-IlePheArgValAlaIleAlaGlyAlaProValThrLeuTrp---Ile-PheYrAsp 772
Db      |||||
842  TTAATCTTCAAGGTGTGATTTGCTGGGGCCCAAGTCAATTCGGGGGATCTTTCTATGAT 901
Qy      |||||
773  ThrGly 774
Db      |||||
902  ACAGGG 907

```

```

RESULT 9
BI084090 1041 bp mRNA linear EST 20-JUN-2001
LOCUS 602869453f1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5013996 5',
DEFINITION mRNA sequence.
ACCESSION BI084090
VERSION BI084090.1 GI:14502420
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1041)
NIH-MGC http://mgi.mci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaags-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.

```

```

FEATURES
    source
        /organism="Homo sapiens"
        /mol_type="RNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:5013996"
        /tissue_type="epidermoid carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH MGC 102"
        /note="Torgan: salivary gland, Vector: pOT87; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCACGAG(G). Library constructed
        by Ling Hong in the laboratory of Gerald M. Rubin
        (University of California, Berkeley) using ZAP-cDNA
        synthesis kit (Stratagene) and Superscript II RT (Life
        Technologies). Note: this is a NIH-MGC Library."
BASE COUNT 280 a 227 c 242 g 292 t
ORIGIN
Alignment Scores:
Pred. No.: 5.5e-157 Length: 1041
Score: 1505.50 Matches: 311
Percent Similarity: 89.04% Conservative: 6
Best Local Similarity: 87.36% Mismatches: 20
Query Match: 32.03% Indels: 21
DB: 12 Gaps: 3
US-10-070-464-1 (1-882) x BI084090 (1-1041)
Qy      |||||
474  LysYrLysArgSerSerGlyGlyLeuProAlaProSerAspPheLysCyProIleLys 493
Db      |||||
3  AAATATAAAGATCCAGGTGTGGGCTGCTGCTCAAGATTTCAAGTCTCTATCAAA 62
Qy      |||||
494  GlnGluIleAlaIleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAanIle 513
Db      |||||
63  GAGGAGATAGCAATTAACAGGTGTAATGGAAAGTCTTGGCGGATGATCTAATATC 122
Qy      |||||
514  GlnValAspGluValArgArgLeuValYrPheGlnGlyThrLysAspSerProLeuGlu 533
Db      |||||
123  CAAGTGTAGAGTCAAGAGGCTGGATATTTTGAAGGACCAAGACCTCCCTTTAGAG 182
Qy      |||||
534  HisHisLeuYrValValSerYrValAsnProGlyGluValThrArgLeuThrAspArg 553
Db      |||||
183  CATCACTGTAGTGTAGTACGTAAATCTCGAGAGGTGACAAAGCTGACCTGACCT 242
Qy      |||||
554  GlyYrSerHisSerCyCyAlIleSerGlnHisCyAaspPheHeIleSerIyYrSer 573
Db      |||||
243  GGTACTACATCTTCTGTCATCAGTCAGCAGCTGACCTCTTTATATAGTAAGTATAGT 302
Qy      |||||
574  AsnGlnLysAspProHisCyValSerLeuYrYrLysLeuSerSerProGluAspAspPro 593
Db      |||||
303  AACCAAGAAATCCACACTGTGTGTCCTTTACAGCTATCAAGTCTTAAGATGACCA 362
Qy      |||||
594  ThrCyLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp 613
Db      |||||
363  ACTTCAAAACAAAGAAATTTGGGCCACATTTGGATTGACAGGTCTCTTCTGACTAC 422
Qy      |||||
614  TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuYrGlyWet 633
Db      |||||
423  TATACTCTCCGAAATTTCTTTTGAAGTACTACGAGATTAACATTTGATGGAGATG 482
Qy      |||||
634  LeuYrLysProHisAspLeuGlnProGlyLysLysYrProThrValLeuPheIleYr 653
Db      |||||
483  CTTCACAGGCTTCATGATTTACAGCTTGAAGAAATATCTTACGTCTGTCATATAT 542
Qy      |||||
654  GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGly--ValLysYrPheArgL 673

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: LNCM1820 row: 5 column: 13
 High quality sequence stop: 847.
 Location/Qualifiers
 1..1041

Db 543 GGTGGTCCACAGGTGCACTTGGTGAATATCGGTTTAAACGACGATCAAGTATTTCGGCT 602
Qy 673 euasnThrleuAlaSerleuGlyTyrValValValIleApeAmaArgGlySerCysH 693
Db 603 TGAATACCCCTAGCTCTCTAGTATGTGGTTGTAGTATGACCAACAGGGGATCTCTGTC 662
Qy 693 iAArgGlyLeuLysPheGluGlyValAAlaPheLysTyrLysMetGlyGlnIleGlnIleAspA 713
Db 663 ACCGAGGGCTTAAATTTGAAGGCGCTTTAA-TATCAATGGCTCATATA-GAAATTGACG 720
Qy 713 spGlnValGluGlyLeuGln-TyrLeuAlaSerArgTyrAspPheIle-AspLeuAspArg 732
Db 721 ATACGTGAAGAGACCTCCAAATTAATCTAGCTTCGATATGATTAATTCATTGACCTTAAGTCG 780
Qy 732 gValGlyIleLeuGlyTyrSerTyrGly-GlyTyrLeuSerLeu-MetAlaLeuMetGln 751
Db 781 TGTGGGATCTCACGGGTGGTGGTCTATGAGAGATACCTCTCCGAAATGGCATTAATGCGAG 840
Qy 752 ArgSerAsp-IlePheArgValAlaIleAla-GlyAlaProValIThrLeuTPIlePhe- 770
Db 841 AGGTCAAGATACCTTCACAGGTGCAATAGCTGGGGGCCACAGTCTGTGATCTCTCT 900
Qy 771 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAengGluGlnGlyTyr 790
Db 901 TATGATACAGATACACGAGACCTTATATGGTCACTGACCTGAACAGA---TGAACGGATCAT 957
Qy 791 TyrLeuGlySerValAlaMetGlnIleGluLysPhePro----- 803
Db 958 ACTTCAGATCTGTG-----CCTGCAGCGAAATTCCTTGAACATGCTTAATGCTCA 1011
Qy 804 SerGluProAsnArgLeuLeuLeuHis 813
Db 1012 ATGGTTCGCGGCAATTCCTGATACAC 1041
RESULT 10 968 bp mRNA linear EST 15-JUL-2002
LOCUS B0671635
DEFINITION AGENCOURT_8203887 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255645
ACCESSION B0671635
VERSION B0671635.1 GI:21782469
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 968)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabds-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2407 row: a column: 22
High quality sequence stop: 630.
Location/Qualifiers
1..968
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6255645"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="PHIOB (phage-resistant)"
/clone_1ib="NIH_MGC_102"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG (G). Library constructed
by Ling Hong in the Laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-CDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 245 a 228 c 235 g 259 t 1 others

ALIGNMENT SCORES:

Pred. No.:	1,35e-148	Length:	968
Score:	1430.00	Matches:	280
Percent Similarity:	87.73%	Conservative:	6
Best Local Similarity:	85.89%	Mismatches:	29
Query Match:	30,43%	Indels:	11
DB:	13	Gaps:	2

US-10-070-464-1 (1-882) x B0671635 (1-968)

Qy 475 TyrLysArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGlu 494
Db 3 TATTAACGATCCAGTGTGGCTGCTGCTCCAAAGATTTCAAGTCTTAACAAAG 62
Qy 495 GlnIleAlaIleThrSerGlyGluTyrProGluValIleuGlyArgHisGlySerAsnIleGln 514
Db 63 GAGATAGCAATTAACAGTGTGAATGGAAATTTCTTGGCCGATGATCTTAATTC 122
Qy 515 ValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHis 534
Db 123 GTTGAATGAAGTACGAAGAGCTGGATATTTTGAAGGACCAAAAGCTCCCTTAGAGCAT 182
Qy 535 HisLeuTyrValValSerTyrValAsnProGlyGluValIThrArgLeuThrAspArgGly 554
Db 183 CACCTGTAGTGTAGTACGTAATCTCGAGAGGTACAAAGGCTGACTACCTGGC 242
Qy 555 TyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsn 574
Db 243 TACTACATTTCTTGCTGATCAGTCACTGACCTGTACTTTTAATGTAAGTATGACTAAC 302
Qy 575 GlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThr 594
Db 303 CAGAAAGATCCACACGTGTGTCTCCCTTTAACAAGCTATCAAGTCTGAAGATGACCAACT 362
Qy 595 CysLysThrLysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyr 614
Db 363 TCCAAACAAAGAAATTTGGGCCACCAATTTGATTCAGCAGTCTCTCTGCTGACTAT 422
Qy 615 ThrProProGlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeu 634
Db 423 ACTCTCCAGAAATTTCTCTTTTAAGTACTGATTAATGATGATGATGATGATGATGATGAT 482
Qy 635 TyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGly 654
Db 483 TACAAGCCATGATCTACAGCTGGAAGAAATATCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 542
Qy 655 GlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsn 674
Db 543 GGTCTTCAGGTGCAGTGTGTGATATCGGTTAAAGAGTCAAGATATTCGGCTTGAT 602
Qy 675 ThrLeuAlaSerLeuGlyTyrValValValValIleApeAmaArgGlySerCysHisArg 694
Db 603 ACCCTAGCTCTCTAGGTTATGTGTGTAGTATGACCAACAGGGGATCTGTGACCGA 662
Qy 696 GlyLeuLysPheGluGlyValAAlaPheLysTyrLysMetGlyGlnIleGlnIleAspArg 714
Db 663 GGGCTTAATTTGAAGCGCGCTTTAATATATAAATGGGTCAAAATGAATATGACATG 722
Qy 715 ValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg-ValGlu 734
Db 723 GTGGAAGACCTCAATATCTAGCTTCTGATATGATTTTCATTGACTTGAATCGAGTGG 782
Qy 734 TLeuIleGlyTyrSer-Tyr-GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 753
Db 783 CATCAAGCGTGGGCGCCATGGAGGATACCTCTCCCTGATGGATTAATGGCCGAG 842

Qy 754 AspiepheargValAlaIlelela-----GIYAlaProvalThlrleuprlephe 770
Db 843 TCCCAATATCTTCCCGGGGCTTATAGCCTCGGGGAGCCCACTCCCGCC----- 893
Qy 771 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyr 790
Db 894 -----GGGGATACCTCCCAAGCAACCGGAGACCCGGGAGCGTTGCCCGGTGGTGC 947
Qy 791 TyrleuGlySer 794
Db 948 CCCCTCGGTGCC 959

RESULT 11
BX372276/c 910 bp mRNA linear EST 08-MAY-2003
LOCUS BX372276 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens CDNA clone CS0DL005YD02 3-PRIME, mRNA sequence.
ACCESSION BX372276 GI:30448117
VERSION BX372276.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 910)
Li W.B., Gruber.C., Jessee.J. and Polayes.D.
Full-length CDNA libraries and normalization
Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7542.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BA1043ZH06_CS04080_1&cluster=7542.r.
Contact: Feng Liang Email: fliang@lifetech.com URL: <http://fulllength.invitrogen.com/InvitrogenCorporation1600>
Paradey Avenue Genoscope sequence ID: CS0BA1043ZH06_CS04080_1.
location/Qualifiers
1..910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL005YD02"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 251 a 193 c 176 g 288 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 2.62e-148 Length: 910
Score: 1427.00 Matches: 293
Percent Similarity: 96.38% Conservative: 0
Best local Similarity: 96.38% Mismatches: 11
Query Match: 30.36% Indels: 5
DB: 13 Gaps: 0

US-10-070-464-1 (1-882) x BX372276 (1-910)

Qy 305 leuGluThrArgGlyAlaAspSerPheArgTyrProlyThrGlyThrAlaAspProlys 324
Db 907 TTGGAAACAGAGGAGGATTCATTCGCG-TATCTAAACAGGACGCAATCTTAA 849
Qy 325 ValThrPheIysMetSerGluIleMetIleAspAlaGluGlyAlaIleleleAspValIle 344

Db 848 GTCACTTTTAAGATGTCAGAAATATATGATTGATGTCGAAGAAAGATCATATGTCAN 789
Qy 345 AspySGluLeuIleGlnProPheGluIleleupheGluGlyValGluTyrIleAlaArg 364
Db 788 GATAAGGAACCTAATTCACCTTTTGAGATTCATT-GAAGNAGTTGGATATATTCACAGA 731
Qy 365 AlaGlyTPThrProGluGlyIleTyrPheAlaIlePheSerIleleuLeuAspArgSerGlnThr 384
Db 730 GCTGATGACCTCCGAGAGAAATATGCTGTGTCATCTCTATCTGATGATGCTCCAGACT 671
Qy 385 ArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGluAspAspValMet 404
Db 670 CGCCTACAGATAGTGTGATCTCACTGATATTTATTCNCAGTGAAGATGATGTATG 611
Qy 405 GluArgGlnArgLeuIleGluSerValProAspSerValThrProLeuIleIleTyrGlu 424
Db 610 GAAAGGCGAGACCTCATGAGTCACTGCTCATCTGAGAGCCCAATATATCATGAA 551
Qy 425 GluThrThrAspIleTPThrIleAsnIleHisAspIlePheHisValPheProGlnSerHis 444
Db 550 GAAACACAGACATCTGATTAATATCCATGACATCTTCATGTTTCCCAAGTCA 491
Qy 445 GluGluGluIleGluPheIlePheIleAspSerGluCysIleThrGlyPheArgHisIleTyr 464
Db 490 GAAGAGGAATTTGAGTTATTTTGTGCTGATGCAAAACAGCTTCCGTCATTTATAC 431
Qy 465 LysIleThrSerIleleuLeuGluSerIleTyrIleArgSerSerGlyIleuProAla 484
Db 430 AAAATTACATCATTTTAAAGAAAGCAAAATATAAACATCCAGTGGGTGCTCTCT 371
Qy 485 ProSerAspPheIysCysProIleleuGluGluIleAlaIleThrSerGlyGluThrGlu 504
Db 370 CCAAGTGAATTCAGTGCTCTCATCAAAAGAGAGATTAATCCAGTGGTGAATGGAA 311
Qy 505 ValLeuGlyArgHisGlySerAsnIleGlnAlaAspGluValArgArgLeuValTyrPhe 524
Db 310 GTTCTGCGCCGCGATGATCTAATATCAAAATTTATGAAAGTCAGAGGCTGTATATTTT 251
Qy 525 GluGlyThrIysAspSerProLeuGluHisIleuTyrValIleSerTyrValAsnPro 544
Db 250 GAAGGACACAAAGACTCCCTTTAGAGATCACTGTAGTCACTTACCTAATCT 191
Qy 545 GlyIleValIleThrArgLeuThrAspArgIleTyrSerHisSerCysCysIleSerGlnHis 564
Db 190 GGAGAGTGAGACAGGCTGACGCGGTGCTACCTCACATTTGCTGATCATGTCAGCAC 131
Qy 565 CysAspPhePheIleSerIleTyrSerAsnGlnIysAsnProHisCysValSerIleTyr 584
Db 130 TGTGACTTCTTTATAGTATGATATGATACAGAAAGATCCACACTGTGTGCTCTTAC 71
Qy 585 LysLeuSerSerProGluAspAspProThrCysIleTyrIleGluPheTyrAlaThrIle 604
Db 70 AAGCATCAATGCTCGAAGATGACCAAC-TGCAGAAACAAAGAAATTTGGGCCACAT- 13
Qy 605 LeuAspSerAla 608
Db 12 TTGATTCAGCA 1

RESULT 12
BX390898 919 bp mRNA linear EST 13-MAY-2003
LOCUS BX390898 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens CDNA clone CS0DL005YD02 5-PRIME, mRNA sequence.
ACCESSION BX390898
VERSION BX390898.1 GI:30607432
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 919)
Li W.B., Gruber.C., Jessee.J. and Polayes.D.

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: sequef@genoscope.cns.fr Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7542.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG053ZD07_CS05044_1&cluster=7542.r.
Contact: Peng Liang Email: fliang@life.techn.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CS0BAG053ZD07_CS05044_1.
Location/Qualifiers

FEATURES
source
1..919
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL005YD02"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 295 a 173 c 197 g 248 t 6 others
ORIGIN

Alignment Scores:
Pred. No.: 6.03e-145 Length: 919
Score: 1397.00 Matches: 286
Percent Similarity: 96.62% Conservative: 0
Best Local Similarity: 96.62% Mismatches: 9
Query Match: 29.72% Indels: 7
DB: 13 Gaps: 0

US-10-070-464-1 (1-882) x BX390898 (1-919)

155 TyrHisGInglYSeGlyThrPheLeuPheGlnAlaGlySer-GlyIleTyrHisVally 174
15 TATTCACCA-GGNAAGGNGACATTTCTGTTCAAGCCGAGTGGGAATTTATCAGCTAA 73
174 SASpGlyGlyProGInglYpHeThGlnGlnProLeuArgProAsnLeuValGluThrSe 194
74 AGA-CGAGGGCCACAAGAGATTACCAACAACCTTTAAGGCCCAATCTAGTGAAGACTAG 132
194 TCySProAsnIleArgMeCAsPProLyLeuCySProAlaAsPProAsGTPiLeAlaPh 214
133 TTGTCCCAACATACGATGAGATCCAAATTAATGAGCCCTGCTGATCCAGACTGATGCTTT 192
214 eIlleHisSerAsnAepIleTpiLleSerAsnIleValThrArgGluGluArgArgLeuTh 234
193 TATACATAGCAACGATATTGATATCTTAACATCTGTAACCAAGAGAAAGAGACTCAC 252
234 rTyrValHisAsnGluLeuAlaAsnMeCgluGluAspAlaArgSerAlaGlyValAlaTh 254
253 TTATGTGCAACATGAGCTAGCCCAACATGAGAAACATGCGAGTGGAGTGGCTAC 312
254 rPheValLeuGlnGluGluPheAspArgTyrSerGlyTyrTPTTPCySProLyValAlG 274
313 CTGTGTTCTCCAAAGAAATTTGATAGATATCTGCAATGCTGTGTCTCCAAAAGCTGA 372
274 uThrThrProSerGlyGlyLySllleLeuArgIleLeuTyrGluGluAsnAspGluSerG 294
373 AACACACTCCCAAGTGTGTAATTTCTTAAGATTTATATGAGAAAGAAATGATGATCTGA 432
294 uValGluIleIleHisValThrSerProMeCLeuGluThrArgArgAlaAsPSerPheAr 314
433 GGTGGAATTAATTACATGTTACATCCCTAATGTTGAAACAAGAGGCGAGATTCAATCCG 492

QY 314 gTyrProLyThrGlyThrAlaAsnProLyValThrPheLyMeCserGluIleMeC 334
DB 493 TTATCTTAAACAGGTACAGCAAAATCTTAAGTCACTTTTAAGATGACGAATAATGAT 552
QY 334 eAspAlaGluGlyArgIleIleAspValIleAspLyGluLeuGlnProPheGlu 354
DB 553 TGATCTTAAGAGAGATCATAGATATGATATGATAGAACTTAATCAACCTTTGAGAT 612
QY 354 eLeuPheGluGlyAlaGlyTyrIleAlaArgAlaGlyTyrThrProGluGlyTyrAl 374
DB 613 TCTATTTAAGAGAGTGAATATATTGCGACAGCTGAGACTCTGAGGGAATAATGC 672
QY 374 aTyrSerIleLeuLeuAspArgSerGlnThrArgLeuGlnIleValIleSerProG 394
DB 673 TTGTCCTACTTACTAGATCCCTCCCAAGCTGCTTACAGATAGTGTGATCTCACCTGA 732
QY 394 uLeuPheIleProValGluAspAspValMeCgluArgGluArgLeuIleGluSerValPr 414
DB 733 ATTATTATCCACAGAGAAAGATGATGTAAGAAAGGACAGACTCATTTAGTCACTGCC 792
QY 414 aAspSerValThrProLeuIleIleTyrGluGluThrThrAspIleTyrPileAsnIleHis 434
DB 793 TGATCTTGAAGCCACCATTAAT-ATCTATGAAAGA-ACAACAACATCTGATTA-ATCCA 849
QY 434 SASpIlePheHisValPheProGInSerHisGluGluGluIleGlu 449
DB 850 TGACATCTTTTCATGNNTTTCCCAAGT-CACGAAAGAGAAATGAG 894

RESULT 13
B1223892
LOCUS B1223892 853 bp mRNA linear EST 11-JUL-2001
DEFINITION 602941035P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5104113 5',
RNA sequence.
ACCESSION B1223892
VERSION B1223892.1 GI:14677336
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMW at:
http://image.llnl.gov

plate: J1AM1251 row: e column: 10
High quality sequence stop: 718.
location/Qualifiers
1..853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5104113"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_lib="NIH MGC 12"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

FEATURES
source

BASE COUNT 250 a 175 c 187 g 241 t
ORIGIN
Alignment Scores:
Pred. No.: 4.88e-142 Length: 853

QY 538 ValValSerTyrValAsnProGlyValValThrArgLeuThrAspArgGlyTyrSerHis 557
 Db 182 GTTGTAACTACAAATCTCTGGAGAGAAAGCACTGACAGAACTGGTACTCTCAT 241
 QY 558 SerCysCysIleSerGlnHisCysAspPheIleSerTyrSerAsnGlnValAsn 577
 Db 242 GCCTCTGTGACGAGATTGACATGTTTCATGACGATGACGACATGAGAAAGC 301
 QY 578 ProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleThr 597
 Db 302 CCACACTGTGTGTCCTTACCGGCTGACAGACATGAAGATGACGCTCAAGAGACA 361
 QY 598 LysGluPheTrrPalaThrIleLeuAspSerIleArgProLeuProAspTyrThrProPro 617
 Db 362 AAGGAATCTGGGCTACAAATTTGGATTCAGAGGCCCTCTCCGATTACATTCCTCCCA 421
 QY 618 GluIlePheSerPheGluSerThrThrArgPheThrLeuTyrGlyMetLeuTyrIlePro 637
 Db 422 GAAGTCTCTCTTGGAGAGCTCCAGGCTTACACTGATGAGATGATGACAAACCT 481
 QY 638 HisAspLeuGlnProGlyLysIleTyrProThrValIleuPheIleTyrGlyLeuProGln 657
 Db 482 CACATCTGCAACCTGGAAGAAAGTACCCTACTGTGATCTTCATCTATGAGAGCCCTCAG 541
 QY 658 ValGlnLeuValAsnAsnArgPheLeuValIleTyrPheArgLeuAsnThrLeuAla 677
 Db 542 GTGACGCTAGTGAAACATTCGATTAAAGACATCAATATTTCCGATTGACCTTGGCC 601
 QY 678 SerLeuGlyTyrValValValIleAspAsnArgGlySerCysHisArgGlyLeuLys 697
 Db 602 TCTTAACTGACGTTGTTGCTGCTTATTTGACAAACCGGGGCTCTGACACGAGGCTGAG 661
 QY 698 PheGluGlyAlaPheLeuTyrLysMetGlyGlnIleGluIleAspAspGlnValGlu-G1 717
 Db 662 TTTCAGAGAGCCTTAAATATACAAATAGGACAAATAGAAATGATGACCAAGTGAAGG 721
 QY 717 YLeuGlnIleLeuAlaSerArgTyrAspPheIleAspLeuAspArgValIleHisG1 737
 Db 722 TCTGACTACTTGTGGCGCTCAGTATGACTCTCATTTATGGCTTCGCGTGGGATTCAGG 781
 QY 737 YTrpSerTyrGlyGlyTyr---LeuSerLeuMetAlaLeuMetGlnArgSer-AspIleP 756
 Db 782 GTGGCTCATGGAGGCTAACTCTCTTATGGCGCTTTTAATGCCGAGGTACAGTACT 841
 QY 756 HeaArgValAlaIleAlaGlyAla-ProValThrLeuTrp--IlePheTyrAspThrGly 774
 Db 842 TCAGGGTTCATCCCGCATCACCCGGTAAACGCTGGGGAGATTCTATAGACACGGGT 901
 QY 775 -TyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySe 794
 Db 902 ATACACGAGAGGCTTAATGGGCCCCCGGGGACACAGAGGAGTATAA-CTGGGTCA 959
 QY 794 rValAlaMetGlnAlaGluLysPhe 802
 Db 960 G---GGCATGCAGAGCGAAGTTTC 981
 RESULT 15
 AL040398 753 bp mRNA linear EST 29-FEB-2000
 LOCUS DKFZp344A0714.r1.434 (synonym: hces3) Homo sapiens cDNA clone
 DEFINITION DKFZp344A0714.5, mRNA sequence.
 ACCESSION AL040398
 VERSION AL040398.1 GI:5409350
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 753)
 AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 TITLE EST (Koehrer, et al.)
 JOURNAL Unpublished
 COMMENT Contact: Koehrer K

MIPS
 Ingolstaedter Landstr.1, D-85764 Neuberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
 Sequenced by BMFZ (Biomedical Research Center at the Charité,
 Berlin/Germany) within the cDNA sequencing consortium of the German
 Genome Project.
 No sl sequence available.
 This clone (DKFZp344A0714) is available at the RZPD in Berlin.
 Please contact the RZPD: Resourcentrum, Heubnerweg 6, 10599
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 source
 1..753
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp344A0714"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_id="434 (synonym: hces3)"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
 BASE COUNT 218 a 150 c 170 g 214 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,68e-139 Length: 753
 Score: 1343.00 Matches: 249
 Percent Similarity: 99.60% Conservative: 0
 Best Local Similarity: 99.60% Mismatches: 1
 Query Match: 28.57% Indels: 0
 DB: 9 Gaps: 0
 US-10-070-464-1 (1-882) x AL040398 (1-753)
 QY 466 IleThrSerIleLeuValGluSerIleTyrIleArgSerArgGlyLeuProAlaPro 485
 Db 1 ATTACATCTATTTTAAAGAAAGCAAAATTAAGATCACTAGTGGGCTGCTCTCCA 60
 QY 486 SerAspPheLysCysProGlyLeuGluGlnIleAlaIleThrSerGlyGluTrrGluVal 505
 Db 60 AGTGAATTCACAGTCTCTCTTAAAGAGAGATGACATTCACAGTGTGAATGGAGATT 120
 QY 506 LeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPheGlu 525
 Db 121 CTGGCGCGCATGATCAATATATCAAGTATGATGACAGAGCTGTATATTGAA 180
 QY 526 GlyThrLysAspSerProLeuGlnHisIleLeuTyrValValSerTyrValAsnProGly 545
 Db 181 GGCACCAAGACTCCCTTTAGACATCACCTGTACGTAGTCAGTACGTAAATCTTGA 240
 QY 546 GluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGlnHisCys 565
 Db 241 GAGGTACAGAGCTGACCTGAGCCGTGGCTACTCACATTTCTCTGATACATGACACACTGT 300
 QY 566 AspPhePheIleSerIleTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLys 585
 Db 301 GACTTCTTATTAAGATAGTATGATACAGAAAGATCACACTGTGTCTCTTTGAAAGTACT 360
 QY 586 LeuSerSerProGluAspAspProThrCysIleThrLysGluPheTrrPalaThrIleLeu 605
 Db 361 CTATCAAGTCTGAAAGATGACCAACTTGCACAAACAAAGAAATTTGGGCACCACTTTTG 420
 QY 606 AspSerIleGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSerThr 625
 Db 421 GATTACAGAGTCTCTTCTCGATATATCTCTCCACAAATTTCTCTTTGAAAGTACT 480
 QY 626 ThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysIle 645
 Db 481 ACTGATTTTCATTTGATGAGATGCTCTACAGAGCTCATGATCTACAGCCCTGGAAGAAA 540
 QY 646 TyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPhe 665

```
Db      541 TATCCTACTGTGCTGTTCAATATATGATGGTCCACAGTGCAGTTGGTGAATATACGTTT 600
Qy      666 LysGlyValIlyTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValVal 685
Db      601 AAAGGAGTCAAGATATTCGGCTTGAAATACCCCTAGCCTCTCTAGATTATGTTGTAGTG 660
Qy      686 IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlnGlyValaPheLysTyrLys 705
Db      661 ATAGACAAACAGGGGATCCTGTCACCGANGGCTTAATTTGAAGGCGCCTTTAATATATAA 720
Qy      706 MetGlyGlnIleGlnIleAspAspGlnVal 715
Db      721 ATGGGTCAATAGAAATTGACGATCAGTG 750
```

Search completed: October 16, 2003, 03:24:56
Job time : 5362 secs

DR N-PSDB; AAC85695.
 XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
 PT substrates, identifying inhibitors of DPP8 catalytic activity which
 PT have therapeutic uses, and for detecting activated T cells -
 XX
 XX Claim 5; Page 68-69; 78pp; English.
 XX
 CC The sequences given in AAB47188-90 represent fragments of human
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for
 CC H-Ala-Pro-pNa, H-Gly-Pro-pNa and H-Arg-Pro-pNa. Therefore, it is a
 CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable
 CC of hydrolysing the peptide bond C-terminal to pro in each of these
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for
 CC cleaving a substrate, and for detecting an activated T cell which
 CC involves measuring the level of DPP8 gene expression in a T cell. The
 CC level of DPP8 expression is detected by detecting the amount of DPP8
 CC RNA in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.
 XX
 XX Sequence 310 AA:
 SQ
 Query Match 100.0%; Score 1680; DB 22; Length 310;
 Best Local Similarity 100.0%; Pred. No. 8.7e-170;
 Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FEGRKDSPLKHLHYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60
 DB 1 FEGRKDSPLKHLHYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60
 QY 61 YKLSPEDDPCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQ 120
 DB 61 YKLSPEDDPCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQ 120
 QY 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYPDIDLRVGIHGMVSGYLSLMAIMOR 180
 DB 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYPDIDLRVGIHGMVSGYLSLMAIMOR 180
 QY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMQAEKFPSEPRLLLL 240
 DB 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMQAEKFPSEPRLLLL 240
 QY 241 HGFIDENHFAHTSILSLFLVRAGKPYDQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300
 DB 241 HGFIDENHFAHTSILSLFLVRAGKPYDQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300
 QY 301 GSRIALAKVI 310
 DB 301 GSRIALAKVI 310

RESULT 2
 AAB08994
 ID ABB08994 standard; protein; 310 AA.
 XX
 AC ABB08994;
 XX
 DT 19-JUN-2002 (first entry)
 XX
 DE Human dipeptidyl peptidase 8.
 XX
 KW Human; dipeptidyl peptidase 8; antiasthmatic; antiallergic;
 KW antiinflammatory.
 XX
 OS Homo sapiens.
 XX
 PN US6337069-B1.
 XX

PD 08-JAN-2002.
 XX
 XX 28-FEB-2001; 2001US-0794236.
 XX
 XX 28-FEB-2001; 2001US-0794236.
 XX
 XX (BMRA-) BMRA CORP BV.
 XX
 XX Grouzmann E, Lacroix J, Monod M;
 XX WPI; 2002-163235/21.
 DR
 PT Treating a patient for mucosal inflammation associated with rhinitis,
 PT sinusitis or both, by intranasally administering a peptidase that
 PT cleaves at Xaa-Pro sequences, to the patient -
 XX
 XX Disclosure; Column 23-24; 13pp; English.
 XX
 CC This invention relates to the treating of a patient for mucosal
 CC inflammation associated with rhinitis or sinusitis, comprising
 CC intranasally administering a peptidase. The peptidase is considered
 CC antiasthmatic, antiallergic and antiinflammatory in its action.
 CC The peptidase cleaves at Xaa-Pro sequences and is useful for treating
 CC a patient for mucosal inflammation associated with rhinitis or
 CC sinusitis, which is the result of allergies or asthma. This
 CC sequence represents human dipeptidyl peptidase 8.
 XX
 XX Sequence 310 AA:
 SQ

Query Match 100.0%; Score 1680; DB 23; Length 310;
 Best Local Similarity 100.0%; Pred. No. 8.7e-170;
 Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FEGRKDSPLKHLHYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60
 DB 1 FEGRKDSPLKHLHYVSYVNGEYTRLTDRGYSHCSCISQCHDFISKYSNOKNPHCVSL 60
 QY 61 YKLSPEDDPCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQ 120
 DB 61 YKLSPEDDPCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQ 120
 QY 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYPDIDLRVGIHGMVSGYLSLMAIMOR 180
 DB 121 KKYPTVLFYGGPQOQIEIDQVEGLQYLASRYPDIDLRVGIHGMVSGYLSLMAIMOR 180
 QY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMQAEKFPSEPRLLLL 240
 DB 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLLSVAMQAEKFPSEPRLLLL 240
 QY 241 HGFIDENHFAHTSILSLFLVRAGKPYDQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300
 DB 241 HGFIDENHFAHTSILSLFLVRAGKPYDQIYPOERHSIRVPESEGEHYELHLHYLOENTL 300
 QY 301 GSRIALAKVI 310
 DB 301 GSRIALAKVI 310

RESULT 3
 AAB47187
 ID AAB47187 standard; Protein; 882 AA.
 XX
 AC AAB47187;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 DE Human DPP8.
 XX
 KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
 KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
 KW growth hormone deficiency; glucose level; mucosal regeneration;
 KW non-insulin dependent diabetes mellitus; glucose intolerance;
 KW immunosuppression.
 KW


```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
FT Active-site 739
FT Active-site 817
FT Active-site /note= "Forms part of Ser-Asp-His catalytic triad"
FT Active-site 849
FT /note= "Forms part of Ser-Asp-His catalytic triad"
XX PN MO200119866-A1.
XX PD 22-MAR-2001.
XX PF 11-SEP-2000; 2000MO-AU01085.
XX PR 10-SEP-1999; 99AU-0002762.
XX PR 18-FEB-2000; 2000AU-0005709.
XX PA (UNSV ) UNIV SYDNEY.
XX PI Abbott CA, Gorell MD;
XX DR WPI; 2001-281520/29.
XX DR N-PSDB; AAC85694.
XX PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
XX PT substrates, identifying inhibitors of DPP8 catalytic activity which
XX PT have therapeutic uses, and for detecting activated T cells -
XX PS Claim 1; Fig 2; 78pp; English.
XX CC This sequence represents human dipeptidyl aminopeptidase (DPP8).
XX CC DPP8 has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and
XX CC H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a
XX CC dipeptidyl peptidase, because it is capable of hydrolysing the
XX CC peptide bond C-terminal to Pro in each of these compounds. DPP8
XX CC is homologous with human DPPIV. DPP8 is useful for cleaving a
XX CC substrate, and for detecting an activated T cell which involves
XX CC measuring the level of DPP8 gene expression in a T cell. The level
XX CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
XX CC in the cell. It is also useful for identifying a molecule capable
XX CC of inhibiting the cleavage of the substrate by DPP8. Molecules
XX CC identified as inhibiting DPP8 catalytic activity may be useful for
XX CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
XX CC in non-insulin dependent diabetes mellitus and other disorders
XX CC involving glucose intolerance, enhancing mucosal regeneration and
XX CC as immunosuppressants.
XX SQ Sequence 882 AA;
Query Match 97.9%; Score 1645.5; DB 22; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
QY 1 FEETKDSPLFHLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISKYSNQKNPHCVSL 60
DB 524 FEETKDSPLFHLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISKYSNQKNPHCVSL 583
QY 61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKXHDLOPG 120
DB 584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKXHDLOPG 643
QY 121 KKPTVLFITGGPQ----- 134
DB 644 KKPTVLFITGGPQVOLVNNRFKGVKXFRNLTLASLGVVVVINDRGSCHRGKLFEGAFK 703
QY 135 ---GOIIDQVVEGLQYLAIRYDFIDLRVGIHGMSYGGYLSLMAIMQSRSDIFRVAIAGA 191
DB 704 YKKGQIIDQVVEGLQYLAIRYDFIDLRVGIHGMSYGGYLSLMAIMQSRSDIFRVAIAGA 763
QY 192 PVTLMIFYDTGYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFIDENVAFA 251

```

```

DB 764 PVTLMIFYDTGYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFIDENVAFA 823
QY 252 HTSILSLFVLRAGKRPYDQIYPOERHSIRVSESGEHHYTLHLYQENUGSRITALKYI 310
DB 824 HTSILSLFVLRAGKRPYDQIYPOERHSIRVSESGEHHYTLHLYQENUGSRITALKYI 882
RESULT 4
ID AAE24170
ID AAE24170 standard; Protein; 882 AA.
XX AC AAE24170;
XX AC AAE24170;
XX DT 23-SEP-2002 (first entry)
XX DE Human dipeptidyl peptidase 8 (DPP8) protein.
XX KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
XX KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
XX KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
XX KW antiviral; enzyme.
XX OS Homo sapiens.
XX PN MO200234900-A1.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001MO-AU01388.
XX PR 27-OCT-2000; 2000AU-0001078.
XX PA (UNSV ) UNIV SYDNEY.
XX PI Abbott CA, Gorell MD;
XX DR WPI; 2002-454646/48.
XX DR N-PSDB; AAD38956.
XX PT New dipeptidyl peptidase (DPP) peptidase, useful for screening
XX PT inhibitors of DPP catalytic activity, which may be employed to treat
XX PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX PT rejection and HIV infection -
XX PS Example; Fig 1; 91pp; English.
XX CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX CC polynucleotides encoding such proteins. The DPP peptidase are useful for
XX CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX CC rejection and HIV (human immuno deficiency virus) infection. The present
XX CC sequence is human DPP8 protein.
XX SQ Sequence 882 AA;
Query Match 97.9%; Score 1645.5; DB 23; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
QY 1 FEETKDSPLFHLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISKYSNQKNPHCVSL 60
DB 524 FEETKDSPLFHLVSVVNPGEVTRLTDRGYSHSCCISOHCDPFISKYSNQKNPHCVSL 583
QY 61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKXHDLOPG 120
DB 584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIRFSFSTTGFTLYGMLYKXHDLOPG 643
QY 121 KKPTVLFITGGPQ----- 134
DB 644 KKPTVLFITGGPQVOLVNNRFKGVKXFRNLTLASLGVVVVINDRGSCHRGKLFEGAFK 703
QY 135 ---GOIIDQVVEGLQYLAIRYDFIDLRVGIHGMSYGGYLSLMAIMQSRSDIFRVAIAGA.191

```

```

Db      704 YKMQIETDQVEGLQYLASRYDFIDLRVGIHMSYCGYLSLWALMQRSDIFRAIAGA 763
Qy      192 PVTLMIFDYGTYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFLENVFA 251
Db      764 PVTLMIFDYGTYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFLENVFA 823
Qy      252 HTSILSLFLVAKGKPYDLQIYPOERHSIRVPESGEHVELHLHYLOENLGSRIALAKVI 310
Db      824 HTSILSLFLVAKGKPYDLQIYPOERHSIRVPESGEHVELHLHYLOENLGSRIALAKVI 882

RESULT 5
ABG61591
ID      ABG61591 standard; Protein; 882 AA.
XX
AC      ABG61591;
XX
DT      12-AUG-2002 (first entry)
XX
DE      Human DPPIV related serine protease DPP-1.
XX
KW      Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KW      DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW      diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW      heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW      ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW      dyskinesia; reproductive disorder; inflammatory disorder;
KW      metabolic disorder.
XX
OS      Homo sapiens.
XX
PN      WO200231134-A2.
XX
PD      18-APR-2002.
XX
PF      12-CT-2001; 2001WO-US31874.
XX
PR      12-OCT-2000; 2000US-240117P.
XX
PA      (FERR) FERRING BV.
XX
PI      Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR      WPI; 2002-444178/47.
DR      N-PSDB; ABK83322.
XX
PT      New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT      the proteins; useful for treating e.g. fungal, bacterial, protozoan and
PT      viral infections; cancers, allergies, neurological disorders, or pain
XX
PS      Claim 17; Fig 1; 113pp; English.
XX
CC      The present invention relates to the isolation of novel human serine
CC      proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC      proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
CC      and nucleic acids encoding them are useful for treating infections
CC      such as fungal, bacterial, protozoan and viral infections, particularly
CC      infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC      pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC      bulimia, Parkinson's disease, acute heart failure, hypotension,
CC      hypertension, urinary retention, osteoporosis, angina pectoris,
CC      stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC      psychotic and neurological disorders (e.g. anxiety, dementia, or
CC      schizophrenia), and dyskinesias. These may also be used in discovering
CC      therapeutic agents for the treatment of reproductive, inflammatory and
CC      metabolic disorders. ABG61591-ABG61612 represent human DPP proteins.
XX
SQ      Sequence 882 AA;
XX
Query Match 97.9%; Score 1645.5; DB 23; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;

```

```

Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
Qy      1 FEETKDSPLREHLVYVYVNGEYTRLTDRGYSHSCCISOHCDFITSKYSNQKPHCVSL 60
Db      524 FEETKDSPLREHLVYVYVNGEYTRLTDRGYSHSCCISOHCDFITSKYSNQKPHCVSL 583
Qy      61 YKLSPEDDPTCKTKEFWATITLDSAGPLPDYTPPELIFSFEFTTGFTLYGMLYKPHDLOPG 120
Db      584 YKLSPEDDPTCKTKEFWATITLDSAGPLPDYTPPELIFSFEFTTGFTLYGMLYKPHDLOPG 643
Qy      121 KKYPVLFYVGGPQ----- 134
Db      644 KKYPVLFYVGGPQVOLLVNNRFKGVKFRLLNTLASLGIVVVVINDRSGCHRGKLFEGAFK 703
Qy      135 ---GQIETDQVEGLQYLASRYDFIDLRVGIHMSYCGYLSLWALMQRSDIFRAIAGA 191
Db      704 YKMQIETDQVEGLQYLASRYDFIDLRVGIHMSYCGYLSLWALMQRSDIFRAIAGA 763
Qy      192 PVTLMIFDYGTYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFLENVFA 251
Db      764 PVTLMIFDYGTYTERYMGHPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFLENVFA 823
Qy      252 HTSILSLFLVAKGKPYDLQIYPOERHSIRVPESGEHVELHLHYLOENLGSRIALAKVI 310
Db      824 HTSILSLFLVAKGKPYDLQIYPOERHSIRVPESGEHVELHLHYLOENLGSRIALAKVI 882

RESULT 6
AAU74749
ID      AAU74749 standard; Protein; 882 AA.
XX
AC      AAU74749;
XX
DT      09-APR-2002 (first entry)
XX
DE      Human protease PRTS-9 protein sequence.
XX
KW      Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW      cardiovascular; atherosclerosis; autoimmune disease; dermatitis;
KW      inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW      cell proliferative disorder; developmental disorder; epilepsy;
KW      Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW      reproductive disorder; endometriosis.
XX
OS      Homo sapiens.
XX
PN      WO200198468-A2.
XX
PD      27-DEC-2001.
XX
PF      13-JUN-2001; 2001WO-US19178.
XX
PR      16-JUN-2000; 2000US-212336P.
PR      22-JUN-2000; 2000US-213955P.
PR      29-JUN-2000; 2000US-215396P.
PR      07-JUL-2000; 2000US-216821P.
PR      14-JUL-2000; 2000US-218946P.
XX
PA      (INCY-) INCYTE GENOMICS INC.
XX
PI      Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM,
PI      Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA,
PI      Walla NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT,
PI      Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L,
PI      Kallik DA;
XX
DR      WPI; 2002-090437/12.
DR      N-PSDB; ABK12892.
XX
PT      Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT      in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT      gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT      proliferative (e.g. cancer) disorders -

```

XX Claim 1; Page 140-142; 177pp; English.
XX
XX The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRTS-9 protein of the invention.
XX
XX Sequence 882 AA;
SQ
Query Match 97.9%; Score 1645.5; DB 23; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
QY 1 FEETKDSPLSEHLLVYVSVNPGVETRLTDRGYSHCCTISQHCDFPISKYSNQKNPHCVSL 60
DB 524 FEETKDSPLSEHLLVYVSVNPGVETRLTDRGYSHCCTISQHCDFPISKYSNQKNPHCVSL 583
QY 61 YKLSPEDDPTCKTEFWATILDSAGPLPDYTPPEIFSESTGTLYGMLYKPHDLQPG 120
DB 584 YKLSPEDDPTCKTEFWATILDSAGPLPDYTPPEIFSESTGTLYGMLYKPHDLQPG 643
QY 121 KKYPVTLFIYGGPO----- 134
DB 644 KKYPVTLFIYGGPOVLVNNRFGVKYFRNLNTLASLGVVVVIDNRGSHGKKEGAFK 703
QY 135 ---GQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRVAIAGA 191
DB 704 YKMQCQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRVAIAGA 763
QY 192 PVTLMIFPDGTYGRYMGHPDQNEGGYVLGSVAQAEPSEPNRLLHGFLENVHFA 251
DB 764 PVTLMIFPDGTYGRYMGHPDQNEGGYVLGSVAQAEPSEPNRLLHGFLENVHFA 823
QY 252 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESGEHLEHLHLYQENLGSRIALAKYI 310
DB 824 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESGEHLEHLHLYQENLGSRIALAKYI 882
RESULT 7
AAG78415
ID AAG78415 standard; Protein: 882 AA.
XX
AC AAG78415;
XX
DT 12-APR-2002 (first entry)
XX
DE Amino acid sequence of 21953 human prolyl oligopeptidase.
XX
KM 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
KM cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KM neuronal disorder; vascular disorder; prostate disorder; cystostatic;
KM antidiabetic; antiautistic; antiautismic; antiinflammatory;
KM diabetes mellitus; arthritis; multiple sclerosis; asthma;
KM Grave's disease; neuronal disorder; demyelinating disease.
XX
OS Homo sapiens.
XX
PN WO200179473-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US40483.
XX

PR 18-APR-2000; 2000US-197508P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX Meyers RA, Williamson M;
PI
XX WPI; 2002-03435/04.
DR
XX N-PsDB; AAH99934.
DR
PT New polypeptides 21953, member of human prolyl oligopeptidase family,
PT useful as diagnostic targets and therapeutic agents for controlling
PT cancer, lymphoma and leukemia
PS
XX Claim 1; Page 102-103; 121pp; English.
XX
XX This invention relates to an isolated 21953 human prolyl
CC oligopeptidase, which is cytosolic, antidiabetic, antiautistic,
CC neuroprotective, antihypertensive, dermatological, antipsoriatic,
CC antiautismic, ophthalmological, antiinflammatory, nootropic,
CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC antitumoral, cardiac, antiatherosclerotic, anorectic and
CC metabolic in its action. Uses include gene therapy, expression or
CC activity of 21953 protein modulator, it is useful for identifying a
CC compound which binds to it and can be used in preventing, treating
CC or detecting a cellular proliferative or differentiative disorder.
CC The 21953 molecules can act as novel diagnostic targets and therapeutic
CC agents for controlling disorders associated with the aberrant activity
CC or degradation of peptide hormones e.g., disorders associated with cell
CC differentiation and proliferation such as cancer, immune function,
CC reproductive, neurological and cardiovascular function. The 21953
CC molecules are thus useful for treating and preventing cellular
CC proliferative and differentiative disorders, hematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the amino
CC acid sequence of 21953 human prolyl oligopeptidase.
XX
SQ Sequence 882 AA;
QY
Query Match 97.9%; Score 1645.5; DB 23; Length 882;
Best Local Similarity 86.4%; Pred. No. 1.9e-165;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
QY 1 FEETKDSPLSEHLLVYVSVNPGVETRLTDRGYSHCCTISQHCDFPISKYSNQKNPHCVSL 60
DB 524 FEETKDSPLSEHLLVYVSVNPGVETRLTDRGYSHCCTISQHCDFPISKYSNQKNPHCVSL 583
QY 61 YKLSPEDDPTCKTEFWATILDSAGPLPDYTPPEIFSESTGTLYGMLYKPHDLQPG 120
DB 584 YKLSPEDDPTCKTEFWATILDSAGPLPDYTPPEIFSESTGTLYGMLYKPHDLQPG 643
QY 121 KKYPVTLFIYGGPO----- 134
DB 644 KKYPVTLFIYGGPOVLVNNRFGVKYFRNLNTLASLGVVVVIDNRGSHGKKEGAFK 703
QY 135 ---GQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRVAIAGA 191
DB 704 YKMQCQIEIDDOVEGLQYLAIRYDFIDLRVGIHGMVSGYSLMALMQRSDIFRVAIAGA 763
QY 192 PVTLMIFPDGTYGRYMGHPDQNEGGYVLGSVAQAEPSEPNRLLHGFLENVHFA 251
DB 764 PVTLMIFPDGTYGRYMGHPDQNEGGYVLGSVAQAEPSEPNRLLHGFLENVHFA 823
QY 252 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESGEHLEHLHLYQENLGSRIALAKYI 310
DB 824 HTSILSLFLVRAKRPYDLQIYPOERHSIRVPESGEHLEHLHLYQENLGSRIALAKYI 882
RESULT 8
ABU07720
ID ABU07720 standard; Protein: 882 AA.
XX

AC ABU07720;
 XX
 DT 19-MAY-2003 (first entry)
 XX
 DE Human serine protease HIPHUM46.
 XX
 KW Human; enzyme; HIPHUM46; serine protease; gene therapy; osteoarthritis;
 KW serine protease activity modulation; dipeptidyl peptidase activity;
 KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KW Alzheimer's disease; paraneoplastic palsy; Huntington's disease;
 KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KW irritable bowel syndrome; type I diabetes; fecal incontinence; anaemia;
 KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KW multiple sclerosis.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Region 259..260
 FT /note="Paired glutamates of the beta propeller domain"
 FT Active-site 739
 FT /label= Catalytic_serine_residue
 FT Active-site 817
 FT /label= Catalytic_aspartate_residue
 FT Active-site 849
 FT /label= Catalytic_histidine_residue
 XX
 PN GB2374869-A.
 XX
 PD 30-OCT-2002.
 XX
 PF 22-JAN-2002; 2002GB-0001404.
 XX
 PR 23-JAN-2001; 2001GB-0001760.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Edbrooke MR, Lewis AP;
 XX
 DR WPI; 2003-150703/15.
 DR N-PSDB; ABX12255.
 XX
 PT Identifying modulators of serine protease activity useful for treating
 PT musculoskeletal diseases, by contacting cell expressing a novel serine
 PT protease polypeptide with a compound and monitoring serine protease
 PT activity -
 XX
 PS Claim 10; Page 26-29; 38pp; English.
 XX
 CC The invention relates to a method of identifying a substance that
 CC modulates serine protease activity, comprising contacting a cell such as
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
 CC virus, expressing a serine protease polypeptide (HIPHUM 46), or its
 CC variant having dipeptidyl peptidase activity, or a serine protease
 CC isolated from the cell with a test substance and monitoring for serine
 CC protease activity. The method is useful for identifying a substance that
 CC modulates serine protease activity. A modulator of the serine protease is
 CC useful in the manufacture of a medicament for treatment or prophylaxis of
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 CC infection, Huntington's disease, paraneoplastic palsy, myotonic
 CC dystrophy, Alzheimer's disease, or amyotrophic lateral sclerosis.
 CC Additional disease that may be treated using modulators of the serine
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung
 CC disease, type I diabetes, fecal incontinence, haemorrhoids, proctitis,
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 CC sclerosis. The present sequence represents the amino acid sequence of the
 CC human serine protease HIPHUM46.
 XX
 SQ Sequence 882 AA;
 Query Match 97.9%; Score 1645.5; DB 24; Length 882;

Best Local Similarity 86.4%; Pred. No. 1.9e-165;
 Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;
 QY 1 FEETKDSPLEHHLYVVSYNVPEVTRLTDRGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 524 FEETKDSPLEHHLYVVSYNVPEVTRLTDRGYSHSCCISQHCDFISKYSNQNPHCVSL 583
 QY 61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYRPEIFSESTTGFTLYGMLYKPHDLQFG 120
 DB 584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYRPEIFSESTTGFTLYGMLYKPHDLQFG 643
 QY 121 KKPPTVLFYTGGRPO----- 134
 DB 644 KKPPTVLFYTGGRPOVQLVNNRFGYKFRFLNLTASIGYVVVVIDNRGSGRGLKFGAIFK 703
 QY 135 ---GQIEIDDOVEGLOYLARSYDFIDLDKVGIGHGNSYGGYLSLMALMQRSDIFRVAIAGA 191
 DB 704 YKMGQIEIDDOVEGLOYLARSYDFIDLDKVGIGHGNSYGGYLSLMALMQRSDIFRVAIAGA 763
 QY 192 PVTLMIFDYDTGYTERYMGHPDNEQGYIGSVAMQAEKPESEPNRLLLHGFIDENVHFA 251
 DB 764 PVTLMIFDYDTGYTERYMGHPDNEQGYIGSVAMQAEKPESEPNRLLLHGFIDENVHFA 823
 QY 252 HTSILSFLVRAGKPYDQIYPOERHSIRVPSGGEHYELHLHYLOENIGSRITAAIKVI 310
 DB 824 HTSILSFLVRAGKPYDQIYPOERHSIRVPSGGEHYELHLHYLOENIGSRITAAIKVI 882
 RESULT 9
 AAB93565
 ID AAB93565 standard; Protein; 632 AA.
 XX
 AC AAB93565;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:12964.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 12964; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 632 AA;

Query Match 83.4%; Score 1401; DB 22; Length 632;
 Best Local Similarity 84.6%; Pred. No. 1.2e-139;
 Matches 269; Conservative 11; Mismatches 20; Indels 18; Gaps 5;

QY 1 FEGTDSPLEHLLYVSVYVNGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 325 FEGTDSPLEHLLYVSVYVNGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 384
 QY 61 YKLSSPEDDPCKTEKFAATILDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQPG 120
 DB 385 YKLSSPEDDPCKTEKFAATILDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQPG 444
 QY 121 KKYPTVLFYGGPQGOIEIDQVEGLQY----LAS-RYDFIDLDRVGI--HGWSYGVYL 172
 DB 445 KKYPTVLFYGGPQGOIEIDQVEGLQY----LAS-RYDFIDLDRVGI--HGWSYGVYL 503
 QY 173 SLMALMQRSDIFRVAIAGAPVTLMIFYDTGYTERYVGHDPDQNEGGYLLGSVAMQAEKPPS 232
 DB 504 KYK-----MVAIAGAPVTLMIFYDTGYTERYVGHDPDQNEGGYLLGSVAMQAEKPPS 554
 QY 233 EPNRLLLHGFIDENVFHATSTILSLFVRAGKPYDLOIYPOBRHSIRVPESGEHYELHL 292
 DB 555 EPNRLLLHGFIDENVFHATSTILSLFVRAGKPYDLOIYPOBRHSIRVPESGEHYELHL 614
 QY 293 LHYLOENLGSRIALAKVI 310
 DB 615 LHYLOENLGSRIALAKVI 632

RESULT 10
 AAB47189 standard; Protein; 465 AA.

XX AAB47189;

XX 29-JUN-2001 (first entry)

DE Human DPP8 318Thr-658Val+759Ala-882Ile.

XX Human, dipeptidyl aminopeptidase; DPP8; prollyl oligopeptidase;
 KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
 KW growth hormone deficiency; glucose level; mucosal regeneration;
 KW non-insulin dependent diabetes mellitus; glucose intolerance;
 KW immunosuppression.

XX Homo sapiens.

XX WO200119866-A1.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000WO-AU01085.

PR 10-SEP-1999; 99AU-0002762.
 PR 18-FEB-2000; 2000AU-0005709.
 XX (UNSY) UNIV SYDNEY.
 PA Abbott CA, Gorell MD;
 PT MPI; 2001-281520/29.
 DR N-PSDB; AAC65696.
 XX
 PS Claim 5; Page 71-72; 78pp; English.

XX The sequences given in AAB47188-90 represent fragments of human
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for
 CC H-Ala-Pro-PNA, H-Gly-Pro-PNA and H-Arg-Pro-PNA. Therefore, it is a
 CC prollyl oligopeptidase and a dipeptidyl peptidase, because it is capable
 CC of hydrolysing the peptide bond C-terminal to Pro in each of these
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for
 CC cleaving a substrate, and for detecting an activated T cell which
 CC involves measuring the level of DPP8 gene expression in a T cell. The
 CC level of DPP8 expression is detected by detecting the amount of DPP8
 CC RNA in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.

XX Sequence 465 AA;

Query Match 82.0%; Score 1377.5; DB 22; Length 465;
 Best Local Similarity 83.5%; Pred. No. 2.4e-137;
 Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 FEGTDSPLEHLLYVSVYVNGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 207 FEGTDSPLEHLLYVSVYVNGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 266
 QY 61 YKLSSPEDDPCKTEKFAATILDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQPG 120
 DB 267 YKLSSPEDDPCKTEKFAATILDSAGPLDPTPEIFSFESTTGFTLYGMLYKPPHDLQPG 326
 QY 121 KKYPTVLFYGGPQGOIEIDQVEGLQYLAIRYDFIDLDRVGIHGWSYGVYLSMALMQR 180
 DB 327 KKYPTVLFYGGPQGOIEIDQVEGLQYLAIRYDFIDLDRVGIHGWSYGVYLSMALMQR 340
 QY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYVGHDPDQNEGGYLLGSVAMQAEKFPSEPNRLLLL 240
 DB 341 -----VAIAGAPVTLMIFYDTGYTERYVGHDPDQNEGGYLLGSVAMQAEKFPSEPNRLLLL 395
 QY 241 HGFIDENVFHATSTILSLFVRAGKPYDLOIYPOBRHSIRVPESGEHYELHLHYLOENL 300
 DB 396 HGFIDENVFHATSTILSLFVRAGKPYDLOIYPOBRHSIRVPESGEHYELHLHYLOENL 455
 QY 301 GSRIALAKVI 310
 DB 456 GSRIALAKVI 465

RESULT 11

ABB97362 standard; Protein; 724 AA.

XX ABB97362;

XX 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 630.

```
XX Human; anti-naemic; vulnerary; anti-inflammatory; immunomodulator;  
KM anti-infectility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; anti-parkinsonian; protein therapy; EST;  
XX expressed sequence tag.  
OS Homo sapiens.  
XX WO200222660-A2.  
XX 21-MAR-2002.  
XX 10-SEP-2001; 2001WO-US26015.  
XX 11-SEP-2000; 2000US-0659671.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
XX N-PSDB; ABN32548.  
XX  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
XX  
XX Example 2; SEQ ID NO 630; 509pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention.  
XX  
XX Sequence 724 AA;  
SQ  
Query Match 82.0%; Score 1377.5; DB 23; Length 724;  
Best Local Similarity 83.5%; Pred. No. 4.6e-137;  
Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;  
OY 1 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDPFISKYSNOKNPHCVSL 60  
DB 466 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDPFISKYSNOKNPHCVSL 525  
OY 61 YKLSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOPG 120  
DB 526 YKLSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOPG 585  
OY 121 KKYPTVLFYGGPGQIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 180  
DB 586 KKYPTVLFYGGPGQIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 599  
OY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 240  
DB 600 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 654  
OY 241 HGFIDENVHFAHTSILLSFVRACKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 300  
DB 655 HGFIDENVHFAHTSILLSFVRACKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 714  
OY 301 GSRIATLKV 310  
DB 715 GSRIATLKV 724
```

RESULT 12
ABB97361

```
ID ABB97361 standard; Protein; 782 AA.  
XX  
XX ABB97361;  
AC  
XX 27-JUN-2002 (first entry)  
XX  
XX Novel human protein SEQ ID NO: 629.  
XX  
XX Human; anti-naemic; vulnerary; anti-inflammatory; immunomodulator;  
KM anti-infectility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; anti-parkinsonian; protein therapy; EST;  
XX expressed sequence tag.  
OS Homo sapiens.  
XX WO200222660-A2.  
XX 21-MAR-2002.  
XX 10-SEP-2001; 2001WO-US26015.  
XX 11-SEP-2000; 2000US-0659671.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
XX N-PSDB; ABN32547.  
XX  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
XX  
XX Example 2; SEQ ID NO 629; 509pp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention.  
XX  
XX Sequence 782 AA;  
SQ  
Query Match 82.0%; Score 1377.5; DB 23; Length 782;  
Best Local Similarity 83.5%; Pred. No. 5.1e-137;  
Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;  
OY 1 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDPFISKYSNOKNPHCVSL 60  
DB 524 FEGTKDSPLEHLLVYVSVNPGEVTRLTDGYSHSCCISOHCDPFISKYSNOKNPHCVSL 583  
OY 61 YKLSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOPG 120  
DB 584 YKLSPEDDPCTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOPG 643  
OY 121 KKYPTVLFYGGPGQIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 180  
DB 644 KKYPTVLFYGGPGQIIEIDQVEGLQYLASRYDFIDLRVGIHGMSTGYLSLMAIMQR 657  
OY 181 SDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 240  
DB 658 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEOGYLLGSVMAQAEKFPSEPNRLILL 712  
OY 241 HGFIDENVHFAHTSILLSFVRACKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 300  
DB 713 HGFIDENVHFAHTSILLSFVRACKPYDLOIYPOERHSIRVPSGEHYELHLHYLOENTL 772
```


QY 301 GSRIALAKVI 310
 DB 773 GSRIALAKVI 782

RESULT 13
 ID AAE14337 standard; Protein; 580 AA.
 AC AAE14337;
 XX
 DT 07-MAR-2002 (first entry)
 DE Human protease PRTS-2 protein.

Human: protease; PRTS-2; tranquilizer; gene therapy; vaccine; allergy; infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis; atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer; gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke; epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea; hypertension; neurological disorder; Parkinson's disease; drug screening; cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis; diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia; autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS; developmental disorder; reproductive disorder; infertility; diarrhoea; dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice.

OS Homo sapiens.
 PN WO200183775-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-US14651.
 XX
 PR 04-MAY-2000; 2000US-202082P.
 PR 11-MAY-2000; 2000US-203586P.
 PR 17-MAY-2000; 2000US-205803P.
 PR 25-MAY-2000; 2000US-207477P.
 PR 01-JUN-2000; 2000US-209402P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.

PI Deleageane AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L,
 PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DAM;
 PI Reddy R, Yue H, Tang YT;
 XX
 DR WPI: 2002-034518/04.
 DR N-PSDB: AAD23843.

PT Novel human proteases and polynucleotides encoding the proteases,
 PT useful for treating, diagnosing or preventing cell proliferative,
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental
 PT disorders -
 PS
 PS Claim 1; Page 120-121; 151pp; English.

The invention relates to human proteases (PRTS1-14) and its corresponding
 CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for
 CC the diagnosis, treatment and prevention of disorders associated with
 CC increased or decreased expression of PRTS. Examples of such disorders
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,
 CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders
 CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic
 CC infections); cardiovascular disorders (myocardial infarction, ischemic
 CC heart disease and hypertension); neurological disorders (epilepsy,
 CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety
 CC and seasonal affective disorder and prion diseases); gastrointestinal

CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder
 CC (infertility, disruption of estrous and menstrual cycle and
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).
 CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic
 CC fragments are useful for screening libraries of compounds in several drug
 CC screening assays. The present sequence is human protease PRTS-2 protein.
 XX

Sequence 580 AA:
 Query Match 81.7%; Score 1372; DB 23; Length 580;
 Best Local Similarity 83.6%; Pred. No. 1,3e-136;
 Matches 266; Conservative 11; Mismatches 19; Indels 22; Gaps 6;

QY 1 FEGTKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 277 FEGTKDSPLEHLLVYVSYVNPGEVTRLTDRGYSHSCCISQHCDFISKYSNQNPHCVSL 336
 QY 61 YKLSPEDDPTCKTKEFMAITLDSAGPLDYPPTPEIFSFESTTGFTLYGMLYKPHDLP 120
 DB 337 YKLSPEDDPTCKTKEFMAITLDSAGPLDYPPTPEIFSFESTTGFTLYGMLYKPHDLP 396
 QY 121 KKYPVTLFIYGGPQCGIETDQVEGLQY-----LAS-RDFTLDLDRVGI--HGMVGYGL 172
 DB 397 KKYPVTLFIYGGPQVQL-VNMRFKGVKFRILTLTSLGIVVVYVIDNRGSGCHRLKEGAF 455
 QY 173 SLMALMQRSDIFRVAIAGAPVTLMIFYDTGTERVYKHDPDQNEQGYLLGSVMAQAEKFP 232
 DB 456 KYK-----MVAIAGAPVTLMIFYDTGTERVYKHDPDQNEQGYLLGSVMAQAEKFP 506
 QY 233 EPNRLLLHGFIDENVHFAHTSILSLFLVRACKPDLQIYPOERHRIKPESEHEHL 292
 DB 507 EPNRLLLHGFIDENVHFAHTSILSLFLVRACKPDLQIYPOERHRIKPESEHEHL 562
 QY 293 LHYIQENLGSRIALAKVI 310
 DB 563 LHYIQENLGSRIALAKVI 580

RESULT 14
 ID AAE24169 standard; Protein; 869 AA.
 XX
 AC AAE24169;
 XX
 DT 23-SEP-2002 (first entry)
 DE Alternating version of murine dipeptidyl peptidase 9 (DPP9) protein.
 XX
 KW Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
 KW human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
 KW type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
 KW antiviral; enzyme.
 XX
 OS Mus sp.
 XX
 PN WO200234900-A1.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-AU01388.
 XX
 PR 27-OCT-2000; 2000AU-0001078.
 PR (UNSY) UNIV SYDNEY.
 PA
 PI Abbott CA, Gorrell MD;
 XX
 DR WPI: 2002-454646/48.
 DR N-PSDB: AAD38955.
 XX

PT New dipeptidyl peptidase (DPP) peptidases, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -

XX Claim 1; Page 70-74; 91pp; English.

XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polynucleotides encoding such proteins. The DPP peptidases are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immuno deficiency virus) infection. The present
 CC sequence is an alternative version of murine DPP9 protein.
 CC Note: This sequence is stated to be the same as that shown as
 CC SEQ ID NO: 4 in the sequence listing of the specification. However these
 CC sequences differ.

XX Sequence 869 AA;

Query Match 65.1%; Score 1094.5; DB 23; Length 869;
 Best Local Similarity 57.3%; Pred. No. 7.5e-107;
 Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 1 FEGTKSPLEHNLVYVSYNPGEVRLTLDRGYSHSCCISQHCDFPISKYSNQKNPCHVSL 60
 DB 521 FQGTNDTPLEHNLVYVSYNPGEVRLTLDRGYSHSCCISQHCDFPISKYSNQKNPCHV 580
 QY 61 YKLSSPEDDPTCKTEFMAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLP 120
 DB 581 YKLSSPEDDPTCKTEFMAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLP 640
 QY 121 KKYPTVLFYGGPO----- 134
 DB 641 RKHPITVLFYGGPOVQLVNNSEFKIKYLRNLTLASLGAVVVIDGRGSCORGLRFEGLAK 700
 QY 135 ---GQIEIDQVYEGLOYLASRYDFIDLRVGHGNSYGGYLSLMALMORSDFRVAIAGA 191
 DB 701 NQMGQVEIDQVYEGLOYLASRYDFIDLRVGHGNSYGGYLSLMALMORSDFRVAIAGA 760
 QY 192 PVTLMIFYDTGYTERMGHPDONEOGYLGVSVAQAEKFPSEPNRLLILHGFLENVHFA 251
 DB 761 PVTVMAYDTGYTERMGHPDONEOGYLGVSVAQAEKFPSEPNRLLILHGFLENVHFA 820
 QY 252 HTSILSPLVRACKPYDLOIYQERHSIRVPSGSEHYEVLHLHYQENL 300
 DB 821 HTNPLVSQILIRACKPYDLOIYQERHSIRVPSGSEHYEVLHLHYQENL 869

RESULT 15

ABG64845 standard; Protein; 497 AA.

XX ABG64845;

DT 27-AUG-2002 (first entry)

XX Human albumin fusion protein #1520.

XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytotoxic; antineoplastic; anti-inflammatory; anticancer;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.

OS Homo sapiens.
 OS Synthetic.

XX WO200177137-A1.

XX 18-OCT-2001.

XX 12-APR-2001; 2001WO-US11988.
 PF 12-APR-2000; 2000US-229358P.
 XX 12-APR-2000; 2000US-199384P.
 PR 25-APR-2000; 2000US-199384P.
 PR 21-DEC-2000; 2000US-256931P.
 XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseeltine WA;

XX WPI; 2002-010886/01.

PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein -

XX Claim 1; Page 1544-1545; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA), also known as human serum
 CC albumin (HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and
 CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
 CC fusion proteins of the invention.

XX Sequence 497 AA;

Query Match 65.1%; Score 1093.5; DB 23; Length 497;
 Best Local Similarity 57.3%; Pred. No. 4.2e-107;
 Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 1 FEGTKSPLEHNLVYVSYNPGEVRLTLDRGYSHSCCISQHCDFPISKYSNQKNPCHVSL 60
 DB 149 FQGTNDTPLEHNLVYVSYNPGEVRLTLDRGYSHSCCISQHCDFPISKYSNQKNPCHV 208
 QY 61 YKLSSPEDDPTCKTEFMAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLP 120
 DB 209 YKLSSPEDDPTCKTEFMAATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLP 268
 QY 121 KKYPTVLFYGGPO----- 134
 DB 269 KKHPTVLFYGGPOVQLVNNSEFKIKYLRNLTLASLGAVVVIDGRGSCORGLRFEGLAK 328
 QY 135 ---GQIEIDQVYEGLOYLASRYDFIDLRVGHGNSYGGYLSLMALMORSDFRVAIAGA 191
 DB 329 NQMGQVEIDQVYEGLOYLASRYDFIDLRVGHGNSYGGYLSLMALMORSDFRVAIAGA 388
 QY 192 PVTLMIFYDTGYTERMGHPDONEOGYLGVSVAQAEKFPSEPNRLLILHGFLENVHFA 251
 DB 389 PVTVMAYDTGYTERMGHPDONEOGYLGVSVAQAEKFPSEPNRLLILHGFLENVHFA 448
 QY 252 HTSILSPLVRACKPYDLOIYQERHSIRVPSGSEHYEVLHLHYQENL 300
 DB 449 HTNPLVSQILIRACKPYDLOIYQERHSIRVPSGSEHYEVLHLHYQENL 497

Search completed: October 15, 2003, 17:10:44
 Job time : 19.9043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 (Search time 6.91621 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1660

Sequence: 1 FEGTKDSPLEHHLVYVSYVN.....HLHYLOENTLGSRIALAKVI 310

Scoring table:

Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1680	100.0	310	4	US-09-794-236-4
2	337.5	20.1	766	4	US-10-002-593-6
3	333.5	19.9	755	5	PCT-US93-07923-3
4	333.5	19.9	759	5	PCT-US93-07923-2
5	333.5	19.9	766	1	US-08-230-491A-3
6	333.5	19.9	766	1	US-08-619-280A-3
7	333.5	19.9	766	2	US-08-940-391-3
8	333.5	19.9	766	4	US-09-794-236-1
9	327	19.5	771	4	US-09-462-284-2
10	302	18.0	760	1	US-08-230-491A-2
11	302	18.0	760	1	US-08-619-280A-2
12	302	18.0	760	2	US-08-940-391-2
13	207	12.3	632	3	US-09-016-080-1
14	152.5	9.1	657	4	US-09-355-166-1
15	147	8.8	721	4	US-09-390-234-20
16	143.5	8.5	622	2	US-08-664-646A-2
17	143.5	8.5	622	2	US-09-066-285-2
18	143.5	8.5	622	2	US-09-261-006-2
19	143.5	8.5	622	3	US-08-951-088-2
20	143.5	8.5	622	4	US-09-609-566-2
21	143.5	8.5	622	4	US-09-609-570-2
22	143.5	8.5	622	4	US-09-427-372-2
23	143.5	8.5	622	4	US-09-693-554-2
24	138	8.2	614	4	US-09-252-991A-20060
25	110	6.5	275	4	US-09-198-452A-181
26	95.5	5.7	344	4	US-09-724-623-79
27	95	5.7	314	4	US-09-355-166-17

28	93.5	5.6	256	4	US-09-355-166-2	Sequence 2, Appl1
29	91.5	5.4	570	3	US-09-068-960-13	Sequence 13, Appl1
30	90.5	5.4	462	4	US-09-134-001C-4300	Sequence 4300, Ap
31	88.5	5.3	255	4	US-09-355-166-3	Sequence 3, Appl1
32	87	5.2	318	4	US-09-601-027-1	Sequence 1, Appl1
33	87	5.2	724	4	US-09-252-991A-20884	Sequence 30884, A
34	85.5	5.1	317	1	US-07-688-299-1	Sequence 1, Appl1
35	85.5	5.1	317	1	US-07-980-517A-1	Sequence 1, Appl1
36	85.5	5.1	318	1	US-07-688-299-3	Sequence 3, Appl1
37	85.5	5.1	318	1	US-07-688-299-13	Sequence 13, Appl1
38	85	5.1	529	4	US-09-252-991A-24711	Sequence 24711, A
39	84	5.0	493	1	US-07-615-448A-7	Sequence 7, Appl1
40	84	5.0	493	1	US-08-196-361-7	Sequence 7, Appl1
41	84	5.0	493	2	US-08-446-934-7	Sequence 7, Appl1
42	84	5.0	493	2	US-08-448-128-7	Sequence 7, Appl1
43	84	5.0	493	3	US-08-948-703-7	Sequence 7, Appl1
44	83.5	5.0	686	3	US-09-368-169-8	Sequence 8, Appl1
45	82	4.9	414	4	US-09-107-532A-5022	Sequence 5022, Ap

ALIGNMENTS

```
RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794.236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-4
```

Query Match 100.0%; Score 1680; DB 4; Length 310;
Best Local Similarity 100.0%; Pred. No. 3.2e-174;
Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FEGTKDSPLEHHLVYVSYVNGEYVRLTDRGSHSCCISOHCDPFIISKYNOXPHCVSL	60
DB	1	FEGTKDSPLEHHLVYVSYVNGEYVRLTDRGSHSCCISOHCDPFIISKYNOXPHCVSL	60
QY	61	YKLSPEDPTCKTEFNATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOG	120
DB	61	YKLSPEDPTCKTEFNATILDSAGPLPDYTPPEIFSFESTTGTLYGMLYKPHDLOG	120
QY	121	KKPYVLFYGGPGQIIDDOVBSGLQYLASRYFDIDRAGIHGWSYGYTSLMALMQR	180
DB	121	KKPYVLFYGGPGQIIDDOVBSGLQYLASRYFDIDRAGIHGWSYGYTSLMALMQR	180
QY	181	SDIRVALAGAPVTLMIYFDYGTERTYVGHDDNOEGYLLSVMAQKFPSEPRLLLL	240
DB	181	SDIRVALAGAPVTLMIYFDYGTERTYVGHDDNOEGYLLSVMAQKFPSEPRLLLL	240
QY	241	HGFLDENVHFATSIILSFLVRACKPYDQIYTPQERHSIRVPESGEHTELHLHYLOENTL	300
DB	241	HGFLDENVHFATSIILSFLVRACKPYDQIYTPQERHSIRVPESGEHTELHLHYLOENTL	300
QY	301	GSRIALAKVI 310	
DB	301	GSRIALAKVI 310	

RESULT 2

```

US-10-002-593-6
; Sequence 6, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTIN
; TITLE OF INVENTION: INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
; FILE REFERENCE: Acty Docket No. 6586198 1242/48/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO: 6
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-593-6

```

Query Match 20.1%; Score 337.5; DB 4; Length 766;

Best Local Similarity 28.0%; Pred. No. 1.3e-27;
Matches 99; Conservative 53; Mismatches 127; Indels 75; Gaps 13;

```

QY 5 KDSPLEHLVYVSYNPGSEVTRLTDRGYSHSCIS-QHCDFFISKYSNOKNPH---C--- 57
DB 423 KMPGGRNLKYIQLIDYTKVTCI-----SCELNPERCQYVSFSEAKYQLRCGSP 475
QY 58 -VSLKLSPEDDPTCKTEFMATILDSAGPLPDYTPPE---IFSFEFTGFTLYGMLYK 113
DB 476 GLPLTLHSSVNDKGLRVLED-NSALDKM--LQNVQMPKSKLDIILNETKF-WYQMIIDP 531
QY 114 PHDLOPKKYPTVLFYIGSP----- 133
DB 532 PH-FPKSKKYPILLDVAAGPCGAKDVFRLNMTATYLASTENIIVASFGRSGYGQDKI 590
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLDRGVIGHWSYGGYLSMALMQRSDIPRV 186
DB 591 MHAIRRLGTFEVEDQIEAARQF-SKMGFVNDKRIALIMGMSYGGYVTSVLGSGGVFKC 649
QY 187 AAGAPVTLTWIIFYDGYTERYMG--HPDQNEQGYLLGSVAMQAEKFPSPENRLLLLHGL 244
DB 650 GIAVAPVSRWEYDVSYTERYMGGLPTPEDNLDRNSTVMSRAENF--KQVEYLLIHGTA 707
QY 245 DENVHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 708 DDNVHFOQSAQISKALVDVGVDFOAMWYTTDEDHGIASSTAHOIYTHMSHFIKQ 761

```

RESULT 3
PCT-US93-07923-3

```

; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schloeman, Stuart F.
; APPLICANT: Tanaka, Yoshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923

```

```

; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Frazer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFO FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3

```

Query Match 19.9%; Score 333.5; DB 5; Length 755;

Best Local Similarity 27.7%; Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;

```

QY 5 KDSPLEHLVYVSYNPGSEVTRLTDRGYSHSCIS-----QHCDFFISKYSNOKNPH--- 56
DB 412 KMPGGRNLKYI-----QLSD-YTKVTCISCELNPERCQYVSFSEAKYQLR 460
QY 57 C---VSLKLSPEDDPTCKTEFMATILDSAGPLPDYTPPE---IFSFEFTGFTLYG 109
DB 461 CSGPLPLTYLHSSVNDKGLRVLED-NSALDKM--LQNVQMPKSKLDIILNETKF-WYQ 516
QY 110 MLKXPHDLOPKKYPTVLFYIGSP----- 133
DB 517 MLPPH-FPKSKKYPILLDVAAGPCGAKDVFRLNMTATYLASTENIIVASFGRSGYGQ 575
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLDRGVIGHWSYGGYLSMALMQRSD 182
DB 576 GDKMHAIRRLGTFEVEDQIEAARQF-SKMGFVNDKRIALIMGMSYGGYVTSVLGSGSG 634
QY 187 AAGAPVTLTWIIFYDGYTERYMG--HPDQNEQGYLLGSVAMQAEKFPSPENRLLLL 240
DB 635 VFKGIAPAIVSRWEYDVSYTERYMGGLPTPEDNLDRNSTVMSRAENF--KQVEYLLI 692
QY 241 HGFLENVHFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 693 HGTADNVHFOQSAQISKALVDVGVDFOAMWYTTDEDHGIASSTAHOIYTHMSHFIKQ 750

```

RESULT 4
PCT-US93-07923-2

```

; Sequence 2, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schloeman, Stuart F.
; APPLICANT: Tanaka, Yoshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX
; SOFTWARE: WordPerfect (Version 5.0)

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Frazer, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 759
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-2

Query Match 19.9%; Score 333.5; DB 5; Length 759;
Best Local Similarity 27.7%; Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;
QY 5 KDSPLEHLLVYVSYNPEBVTRLTDGYSHSCCIS-----QHCDFFISKYSNQKNPH--- 56
DB 416 KMPGGRMLYKRI-----QLSD--YTKVTCLSCELNPERCOYVSFSKAKYQQLR 464
QY 57 C-----VSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPPE---ISFESTTGFTLYG 109
DB 465 CSGPGLPLYTHSSVNDKGLRVLED-NSALDKM--LQNVQWPSKLDPTILNETKF-WYQ 520
QY 110 MLYKPHDLQPKKPYTVLFYIGP----- 133
DB 521 MILPPH-FDKSKKPYLLDDVYAGPSCQKADTVFRLMATYLASTENIIVASFDGSGSQY 579
QY 134 -----QGIEIDQVEGLQYLASRYDFIDLDRVGIHGSYGYLSIMALMORS 182
DB 580 GDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWMSYGGYVSMVLGSSG 638
QY 183 IFRVALIAGAPVTLMFYDGTGTERYMG--HPDQNEQGYLGSVMAQAEKFPSEPRRLLL 240
DB 639 VFKGCIAPVSRMEYDVSITERYMGLPTEPDNLDRNSTVMSRAENF--KQVEYLLI 696
QY 241 HGFLENVHFAHTSILSLFLVRACKPYDLQIYPOERHSIRVPESEHYELHLHYLOE 298
DB 697 HGTADDNVHFGQSAQISKALVDGVDFQAMMYTDEDHGIASTAHQHIYTHMSHFIRK 754

RESULT 5
US-08-230-491A-3
Sequence 3, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND US93
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELFE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-3

Query Match 19.9%; Score 333.5; DB 1; Length 766;
Best Local Similarity 27.7%; Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;
QY 5 KDSPLEHLLVYVSYNPEBVTRLTDGYSHSCCIS-----QHCDFFISKYSNQKNPH--- 56
DB 423 KMPGGRMLYKRI-----QLSD--YTKVTCLSCELNPERCOYVSFSKAKYQQLR 471
QY 57 C-----VSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPPE---ISFESTTGFTLYG 109
DB 472 CSGPGLPLYTHSSVNDKGLRVLED-NSALDKM--LQNVQWPSKLDPTILNETKF-WYQ 527
QY 110 MLYKPHDLQPKKPYTVLFYIGP----- 133
DB 528 MILPPH-FDKSKKPYLLDDVYAGPSCQKADTVFRLMATYLASTENIIVASFDGSGSQY 586
QY 134 -----QGIEIDQVEGLQYLASRYDFIDLDRVGIHGSYGYLSIMALMORS 182
DB 587 GDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWMSYGGYVSMVLGSSG 645
QY 183 IFRVALIAGAPVTLMFYDGTGTERYMG--HPDQNEQGYLGSVMAQAEKFPSEPRRLLL 240
DB 646 VFKGCIAPVSRMEYDVSITERYMGLPTEPDNLDRNSTVMSRAENF--KQVEYLLI 703
QY 241 HGFLENVHFAHTSILSLFLVRACKPYDLQIYPOERHSIRVPESEHYELHLHYLOE 298
DB 704 HGTADDNVHFGQSAQISKALVDGVDFQAMMYTDEDHGIASTAHQHIYTHMSHFIRK 761

RESULT 6
US-08-619-280A-3
Sequence 3, Application US/06619280A
Patent No. 5767242
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPFA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-619-280A-3
```

```

Query Match          19.9%; Score 333.5; DB 1; Length 766;
Best Local Similarity 27.7%; Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;
```

```

QY 5 KDSPLEHLYVSYNPGVTRLTDRGYSHSCIS-----QHCDFFISKYSNOKNPH--- 56
DB 423 KMPGGRNLTKI-----QLSD--YTKVTCLSCENLPERCOYVSFSKEANYQLR 471
QY 57 C-----VSLYKLSPEDDPTCKTEFMATILDSAGLPDYTPPE---IFSFEETGFTLYG 109
DB 472 CSGPLPLTYLTHSSVNDKGLRVLED-NSALDKM--LQNVQPSKLDLIINETRF-WYQ 527
QY 110 MLKXPHDLPQKKYPTVLFYIGP----- 133
DB 528 MLPPH-FDKSKKYPLLLDVYAGPCSQKADTVFRLNMTYLASTENIIVASFQSGSGQ 586
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLRVGHGWSYGYLSMALMQRSD 182
DB 587 GDKIMHAINRLRGTEVEVDQIEAARQF-SKMGFVNDKRIAIWMSYGYVTSMLVSGSG 645
QY 183 IFRVAIAGAPVTLMIFYDTGYTERYMG--HPDQNEGYGLSGVAMQAEKFPSEPRRLLL 240
DB 646 VFKCGIAVAPSRMEYDVSYTERYMGLPTEPDNLDHYNSTVMSRAENF--KQVEYLLI 703
QY 241 HGFLENNHFAHTSILSLFLVAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 704 HGTADNVHFOOSAISKALVDVGVDFOAMWYTTDEDHGIASSTAHQHYTHMSHFIRKQ 761
```

```

RESULT 7
US-08-940-391-3
Sequence 3, Application US/08940391
Patent No. 5965373
```

```

GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280
FILING DATE: 18-MARCH-1996
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5965373man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-940-391-3
```

```

Query Match          19.9%; Score 333.5; DB 2; Length 766;
Best Local Similarity 27.7%; Pred. No. 3.5e-27;
Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;
```

```

QY 5 KDSPLEHLYVSYNPGVTRLTDRGYSHSCIS-----QHCDFFISKYSNOKNPH--- 56
DB 423 KMPGGRNLTKI-----QLSD--YTKVTCLSCENLPERCOYVSFSKEANYQLR 471
QY 57 C-----VSLYKLSPEDDPTCKTEFMATILDSAGLPDYTPPE---IFSFEETGFTLYG 109
DB 472 CSGPLPLTYLTHSSVNDKGLRVLED-NSALDKM--LQNVQPSKLDLIINETRF-WYQ 527
QY 110 MLKXPHDLPQKKYPTVLFYIGP----- 133
DB 528 MLPPH-FDKSKKYPLLLDVYAGPCSQKADTVFRLNMTYLASTENIIVASFQSGSGQ 586
QY 134 -----QGQIEIDQVEGLQYLASRYDFIDLRVGHGWSYGYLSMALMQRSD 182
DB 587 GDKIMHAINRLRGTEVEVDQIEAARQF-SKMGFVNDKRIAIWMSYGYVTSMLVSGSG 645
QY 183 IFRVAIAGAPVTLMIFYDTGYTERYMG--HPDQNEGYGLSGVAMQAEKFPSEPRRLLL 240
DB 646 VFKCGIAVAPSRMEYDVSYTERYMGLPTEPDNLDHYNSTVMSRAENF--KQVEYLLI 703
QY 241 HGFLENNHFAHTSILSLFLVAGKPYDLOIYPOERHSIRVPESGEHYELHLHYLOE 298
DB 704 HGTADNVHFOOSAISKALVDVGVDFOAMWYTTDEDHGIASSTAHQHYTHMSHFIRKQ 761
```

```

RESULT 8
US-09-794-236-1
Sequence 1, Application US/09794236
Patent No. 6337069
```

```

GENERAL INFORMATION:
APPLICANT: Grouzmann, Eric
APPLICANT: Lacroix, Jean-Silvain
APPLICANT: Monod, Michel
TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
FILE REFERENCE: 81985/276823
CURRENT APPLICATION NUMBER: US/09/794,236
CURRENT FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 766
TYPE: PRT
ORGANISM: Homo sapiens
US-09-794-236-1
```

```

Query Match          19.9%; Score 333.5; DB 4; Length 766;
Best Local Similarity 27.7%; Pred. No. 3.5e-27;
```



```
Db 577 TAFQCDKLLVAVYRKLGVEVEQITAVKFI-EMGFIDEKRIALIMGMEYGGVSLALA 635
Qy 179 QRSDFRVAIAGAPVTLMIFYDTGTERMGHP--DQNEQGYLLGSVAMQAKPSEPRK 236
Db 636 SGTGFKCGIAVAVPSSWEYASVYTERFMGLPTDQDNLEHKNSTVMAAEYFRVND-- 693
Qy 237 LLLHGFLENVFAHTSILLSFLVRAGKPYDQIYPOERHSIRVDESG--EHEYELHL 293
Db 694 YLLHGTADNDVHFQNSAQIALVLAQVDFQAMYSIDQNHGL---SGLSTNHLTYTMT 749
Qy 294 HYLQE 298
Db 750 HFLKQ 754
```

RESULT 13

```
US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1
```

Query Match 12.3%; Score 207; DB 3; Length 632;
Best Local Similarity 25.4%; Pred. No. 1,5e-13;
Matches 61; Conservative 38; Mismatches 79; Indels 62; Gaps 7;

```
Qy 94 PEIPEFESTGTFTLQGMLYKPHDLQPGKKYPTVLFYGGPO----- 134
Db 375 PEHFFVKASDGEVIDAMWKPVNFRRKQKYPALILEIHGSPKTAAGYAFMEHFVLTSGKF 434
Qy 135 -----GOIEIDQVEGLQYLASRYDFIDLRVGIHGMSTYGG 170
Db 435 VVIFSNPRGSDGYGEFADIRGHYGERDYQDLMEVVDLAKRFDFIDGERLDVGTGSGYGG 494
Qy 171 YLSLMAHQRSDFRVAIAGAPVTLMIFY---DTGYTERYMGHPDQ-----NEQGY 219
Db 495 FMTNW-IVGHTNRFKAAVLTQRSISNMWISFPGTTDGY--YFA-PDQICKDPWSNLEGYW 549
Qy 220 LGSVAMQAKPSEPRNLLHGFLENVFAHTSILLSFLVRAGKPYDQIYPOERHSI 279
Db 550 EKSPKLYA---PNVETPLTIHSTEDYRCWLPDALQFLISLKYLDKREVLALFPGENHDL 606
```

RESULT 14

```
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GCS11-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1
```

Query Match 9.1%; Score 152.5; DB 4; Length 657;
Best Local Similarity 20.1%; Pred. No. 1.3e-07;
Matches 72; Conservative 48; Mismatches 121; Indels 117; Gaps 12;

```
Qy 4 TKDSPLEHNLVYVSNPGEVTRLTRDRC-----YSHSCCISQHC 42
Db 311 TKDS---QGFFYIG-----TDQSGTGIYISIEGLVPIRLKEXYINSFSLSPDE 357
Qy 43 DFFISKTSNQKPHCVSLKYLSSPEDDPCTKTEFWATILDSAGLPDYT--PEIPEFE 100
Db 358 OHFIASVTKPDRPSSLYSIPLGOEKKOLTGANDKF-----VREHTISIPERIQYA 407
Qy 101 STTGFTLYGMLYKPHDLQPGKKYPTVLFYGGPO----- 134
Db 408 TEDGVAMVGMWLMRPQMEGETTYPLILNHGGPHMYGHTTHERFOVLAAGYAVVYINP 467
Qy 135 -----GOIEIDQVEGLQYLASRYDFIDLRVGIHGMSTYGYLSLMA 176
Db 468 RSHGYYGGEFVNAVAVGDKQYDDVMQAVDEAIKRDPHIDPKRLGVGTGSGYGFMTNW- 526
Qy 177 LMQRSDIRVAIAGAPVTLMIFY---DTGY--TERYMGHP-----DQNEQGYIYG 221
Db 527 IVGQTNRRKAAVLTQRSISNMWISFHGVSIDIGYFFTWOLEHMFEDTEKLMDSPLKY--- 583
Qy 222 SVAMQAKPSEPRNLLHGFLENVFAHTSILLSFLVRAGKPYDQIYPOERHSI 279
Db 584 --AANVE-----TPLLILHGERDRCPIEQAEQLFIALKKMGKETKLVFPNASHNL 633
```

RESULT 15

```
US-09-390-234-20
; Sequence 20, Application US/09390234
; Patent No. 6365390
; GENERAL INFORMATION:
; APPLICANT: Blum, David L.
; APPLICANT: Kataeva, Irina
; APPLICANT: Li, Xin-liang
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Phenolic Acid Esterases, Coding Sequences and Methods
; FILE REFERENCE: 67-98
; CURRENT APPLICATION NUMBER: US/09/390,234
; CURRENT FILING DATE: 1999-09-03
; EARLIER APPLICATION NUMBER: US 60/099,136
; EARLIER FILING DATE: 1998-09-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-09-390-234-20
```

Query Match 8.8%; Score 147; DB 4; Length 721;
Best Local Similarity 20.6%; Pred. No. 6.1e-07;
Matches 58; Conservative 33; Mismatches 98; Indels 92; Gaps 8;

```
Qy 93 PEIPEFESTGTFTLQGMLYKPHDLQPGKKYPTVLFYGGPO----- 136
Db 439 PSDISEFYQGNFTDIIHAMVIYFENFDKSKYPLFIHGGPQGMADGWSRWMPKAWA 498
Qy 137 -----IEIDQVEGLQYLASRYDFIDLRVGIHGMSTYGYLSLMA 165
Db 489 DQGYVVAAPNPTGSGFGQALTTAIQNNMGAPYDQIYPOERHSI 558
Qy 166 WSYGYL-----SLMALQRSDFRVAIAGAPVT--LWIFYDTGYTERYMGHP 211
```

Db	559	ASVGGFMINWIOSSPLGRKKALVSHDGT	----	VADAKVSTEELWFM	-----	602
Qy	212	DONEQYIYIGSVAMQAEKF	----	PSEPNR	-----	602
Db	603	---QREFNGTWDARDNRRWDPSA	PERILOPATPM	LVHSDKDYRLP	VAGLSL	FNVL 658
Qy	261	VRAGKPYDLOIYPOERHSIRV	PESGHEYLHLHY	LOENIG	301	
Db	659	QERGVP	SRPLNFPDENHWVVPENSL	VWHQALGW	INKYSG	699

Search completed: October 15, 2003, 17:15:35
Job time : 7.91621 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 15, 2003, 17:13:57 ; Search time 43.8027 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680
Sequence: 1 FEGRKDSPLFHLVVSYN.....HLHLYGRLGRIALAKVI 310

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1680	100.0	310	12	US-09-993-959-4
2	1645.5	97.9	882	10	US-09-976-674-1
3	1645.5	97.9	882	12	US-10-054-776-2
4	1645.5	97.9	882	12	US-10-170-789-38
5	1093.5	65.1	863	10	US-09-976-674-3
6	1093.5	65.1	892	10	US-09-976-674-23
7	1093.5	65.1	892	10	US-09-976-674-27
8	1086.5	64.7	879	10	US-09-976-674-3
9	1086.5	64.7	879	10	US-09-976-674-35
10	1032.5	61.5	508	15	US-10-237-271-3
11	739	44.0	658	10	US-09-976-674-19
12	739	44.0	660	10	US-09-976-674-11
13	629.5	37.5	832	10	US-09-976-674-29
14	629.5	37.5	832	10	US-09-976-674-31
15	629.5	37.5	832	10	US-09-976-674-31

16	622.5	37.1	819	10	US-09-976-674-37	Sequence 37, Appl
17	622.5	37.1	819	10	US-09-976-674-39	Sequence 39, Appl
18	582	34.6	129	11	US-09-764-691-3564	Sequence 3564, Ap
19	471	28.0	613	10	US-09-976-674-21	Sequence 21, Appl
20	344.5	20.5	767	15	US-10-165-603-4	Sequence 4, Appl
21	337.5	20.1	766	12	US-10-423-714-6	Sequence 6, Appl
22	337.5	20.1	766	14	US-10-002-593-6	Sequence 6, Appl
23	337.5	20.1	766	15	US-10-165-603-7	Sequence 7, Appl
24	333.5	19.9	766	9	US-09-265-606-3	Sequence 3, Appl
25	333.5	19.9	766	12	US-09-993-959-1	Sequence 3, Appl
26	310	18.5	710	15	US-10-156-761-10681	Sequence 1, Appl
27	310	18.5	818	12	US-10-401-437-3	Sequence 10681, A
28	310	18.5	818	12	US-10-402-312-3	Sequence 3, Appl
29	310	18.5	818	12	US-10-402-667-3	Sequence 3, Appl
30	310	18.5	818	12	US-10-401-436-3	Sequence 3, Appl
31	302	18.0	760	9	US-09-265-606-2	Sequence 2, Appl
32	302	18.0	760	12	US-10-301-822-55	Sequence 55, Appl
33	302	18.0	760	15	US-10-177-293-136	Sequence 136, App
34	277	16.5	706	10	US-09-976-674-41	Sequence 41, Appl
35	277	16.5	796	10	US-09-976-674-5	Sequence 5, Appl
36	277	16.5	796	12	US-09-870-133-2	Sequence 2, Appl
37	277	16.5	796	15	US-10-160-501-5	Sequence 5, Appl
38	267.5	15.9	691	10	US-09-976-674-43	Sequence 43, Appl
39	254	15.1	164	15	US-10-106-698-6740	Sequence 6740, Ap
40	240	14.3	803	12	US-10-205-219-58	Sequence 58, Appl
41	152.5	9.1	657	9	US-09-950-368-1	Sequence 1, Appl
42	152.5	9.1	657	12	US-10-401-437-2	Sequence 2, Appl
43	152.5	9.1	657	12	US-10-402-312-2	Sequence 2, Appl
44	152.5	9.1	657	12	US-10-402-667-2	Sequence 2, Appl
45	152.5	9.1	657	12	US-10-401-436-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1	US-09-993-959-4	Application US/09993959
Sequence 4, Appl	US/09993959	
Publication No.	US20030165489A1	
GENERAL INFORMATION:		
APPLICANT:	Grouzmann, Eric	
APPLICANT:	Lacroix, Jean-Silvain	
APPLICANT:	Monod, Michel	
TITLE OF INVENTION:	Method of Treating Rhinitis and Sinusitis	
FILE REFERENCE:	81985/276823	
CURRENT APPLICATION NUMBER:	US/09/993, 959	
CURRENT FILING DATE:	2001-11-27	
NUMBER OF SEQ ID NOS:	4	
SOFTWARE:	PatentIn version 3.0	
SEQ ID NO 4		
LENGTH:	310	
TYPE:	PRT	
ORGANISM:	Homo sapiens	
US-09-993-959-4		
Query Match	100.0%	Score 1680; DB 12; Length 310;
Best Local Similarity	100.0%	Pred. No. 2.4e-165;
Matches 310;	Conservative 0;	Mismatches 0;
Indels 0;	Gaps 0;	
QY	01	FEGRKDSPLFHLVVSYNPGEVTRLDRGYSHSCISQCHDFISKYSNOKPHCVSL 60
DB	1	FEGRKDSPLFHLVVSYNPGEVTRLDRGYSHSCISQCHDFISKYSNOKPHCVSL 60
QY	61	YKSSPEDDPCKTKEFMATILDSAGPLPDYTPPEIFSFESTGTLYGMLYKPHDQPG 120
DB	61	YKSSPEDDPCKTKEFMATILDSAGPLPDYTPPEIFSFESTGTLYGMLYKPHDQPG 120
QY	121	KKYTVTFYGGPGGQIEIDOVGEGLOYLASRYFILDVRGHWGSGYGLSLMALMOR 180
DB	121	KKYTVTFYGGPGGQIEIDOVGEGLOYLASRYFILDVRGHWGSGYGLSLMALMOR 180
QY	181	SDIFRVAAGAPVTLTFYDGTGYTERYMGHPDQEGGYLGSVMQAEKFFSEENRLLLL 240

Db 181 SDFIRVAVAGAPVTLWIFYDTGTYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLL 240
 QY 241 HGFLENVFAHTSILSLFVRAGKPYDLOIYPOERHSIRVPSGEGHEHLHLYQENL 300
 Db 241 HGFLENVFAHTSILSLFVRAGKPYDLOIYPOERHSIRVPSGEGHEHLHLYQENL 300
 QY 301 GSRIALAKVI 310
 Db 301 GSRIALAKVI 310

RESULT 2
 US-09-976-674-1
 ; Sequence 1, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akimsanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Julien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 882
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-976-674-1

Query Match 97.9%; Score 1645.5; DB 10; Length 882;
 Best Local Similarity 86.4%; Pred. No. 4e-161; Indels 49; Gaps 1;
 Matches 310; Conservative 0; Mismatches 0

QY 1 FEGTKDSPLEHLLYVSVYNPGEVTRLTDRGYSHSCISQHCDFISKYSNQNPHCVSL 60
 Db 524 FEGTKDSPLEHLLYVSVYNPGEVTRLTDRGYSHSCISQHCDFISKYSNQNPHCVSL 583
 QY 61 YKLSSPEDPTCKTEFMAITLDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 120
 Db 584 YKLSSPEDPTCKTEFMAITLDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 643
 QY 121 KKPPTVLFYGGPO----- 134
 Db 644 KKPPTVLFYGGPOVOLVNNRFGVYFRRLNTLASLGIYVVVVVDNRGSGHRLKEGAKR 703
 QY 135 ---GQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYLSMALMORSDFRVAIAGA 191
 Db 704 YKMGQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYLSMALMORSDFRVAIAGA 763
 QY 192 PPTLWIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 251
 Db 764 PPTLWIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 823
 QY 252 HTSILSLFVRAGKPYDLOIYPOERHSIRVPSGEGHEHLHLYQENLGSRIALAKVI 310
 Db 824 HTSILSLFVRAGKPYDLOIYPOERHSIRVPSGEGHEHLHLYQENLGSRIALAKVI 882

RESULT 3
 US-10-054-776-2
 ; Sequence 2, Application US/10054776
 ; Publication No. US20030165818A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mark Robert Edbrooke
 ; APPLICANT: Alan Peter Lewis
 ; TITLE OF INVENTION: NOVEL PROTEIN
 ; FILE REFERENCE: OG1042US
 ; CURRENT APPLICATION NUMBER: US/10/054,776

; CURRENT FILING DATE: 2002-01-23
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 882
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-054-776-2

Query Match 97.9%; Score 1645.5; DB 12; Length 882;
 Best Local Similarity 86.4%; Pred. No. 4e-161; Indels 49; Gaps 1;
 Matches 310; Conservative 0; Mismatches 0

QY 1 FEGTKDSPLEHLLYVSVYNPGEVTRLTDRGYSHSCISQHCDFISKYSNQNPHCVSL 60
 Db 524 FEGTKDSPLEHLLYVSVYNPGEVTRLTDRGYSHSCISQHCDFISKYSNQNPHCVSL 583
 QY 61 YKLSSPEDPTCKTEFMAITLDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 120
 Db 584 YKLSSPEDPTCKTEFMAITLDSAGPLDPYTPPEIFSFESTTGFTLYGMLYKPHDLQNG 643
 QY 121 KKPPTVLFYGGPO----- 134
 Db 644 KKPPTVLFYGGPOVOLVNNRFGVYFRRLNTLASLGIYVVVVVDNRGSGHRLKEGAKR 703
 QY 135 ---GQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYLSMALMORSDFRVAIAGA 191
 Db 704 YKMGQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYLSMALMORSDFRVAIAGA 763
 QY 192 PPTLWIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 251
 Db 764 PPTLWIFYDTGYTERYMGHPDQNEOGYVLSVAMQAEKFPSEPNRLLLHGFLENVFA 823
 QY 252 HTSILSLFVRAGKPYDLOIYPOERHSIRVPSGEGHEHLHLYQENLGSRIALAKVI 310
 Db 824 HTSILSLFVRAGKPYDLOIYPOERHSIRVPSGEGHEHLHLYQENLGSRIALAKVI 882

RESULT 4
 US-10-170-789-38
 ; Sequence 38, Application US/10170789
 ; Publication No. US20030180930A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rachel E. Meyers
 ; APPLICANT: Oladot, Peter J.
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; APPLICANT: Curtis, Rory A. J.
 ; APPLICANT: Williams, Mark
 ; APPLICANT: Welch, Nadine
 ; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
 ; FILE REFERENCE: 10448-191001
 ; CURRENT APPLICATION NUMBER: US/10/170,789
 ; CURRENT FILING DATE: 2002-06-13
 ; PRIOR APPLICATION NUMBER: US 09/797,039
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: PCT/US01/06525
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/186,061
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 09/882,166
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: PCT/US01/19269
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: US 60/212,078
 ; PRIOR FILING DATE: 2000-06-15
 ; PRIOR APPLICATION NUMBER: US 09/934,406
 ; PRIOR FILING DATE: 2001-08-21
 ; PRIOR APPLICATION NUMBER: PCT/US01/26052
 ; PRIOR FILING DATE: 2001-08-21
 ; PRIOR APPLICATION NUMBER: US 60/226,740
 ; PRIOR FILING DATE: 2000-08-21
 ; PRIOR APPLICATION NUMBER: US 09/861,801


```
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: PCT/US01/16549
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: US 60/205,508
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/801,267
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07138
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,454
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/829,671
/ PRIOR FILING DATE: 2001-04-10
/ PRIOR APPLICATION NUMBER: PCT/US01/40483
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: US 60/197,508
/ PRIOR FILING DATE: 2000-04-18
/ PRIOR APPLICATION NUMBER: US 09/961,721
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: PCT/US01/29904
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: US 60/235,023
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US 10/045,367
/ PRIOR FILING DATE: 2001-11-07
/ PRIOR APPLICATION NUMBER: US 60/246,561
/ PRIOR FILING DATE: 2000-11-07
/ PRIOR APPLICATION NUMBER: US 09/801,275
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07074
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,420
/ PRIOR FILING DATE: 2000-03-07
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 38
/ LENGTH: 882
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-170-789-38
```

```
Query Match      97.9%; Score 1645.5; DB 12; Length 882;
Best Local Similarity 86.4%; Pred. No. 4e-161;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

QY      1  FEGTDSPLEHLLVYVSVNPEGEVRLTDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      524  FEGTDSPLEHLLVYVSVNPEGEVRLTDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 583

QY      61  YLSSPEDDPCTCKTEFNATLDSAGPLPDYTPPEIFSFESTTGTGMLTKPHDLOG 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      584  YLSSPEDDPCTCKTEFNATLDSAGPLPDYTPPEIFSFESTTGTGMLTKPHDLOG 643

QY      121  KKYPTLVFIYGGPQ----- 134
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      644  KKYPTLVFIYGGPQVQLVNNRFKGVKVFRLNTLASLGYVVVVIDNRGSCHRGLKEGAFK 703

QY      135  ---GOIEIDDOVEGIQYLASRYDFIDLDRVGIHGSYGGYSLMALMORSDFRVAIAGA 191
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      704  YKMGIEIDDOVEGIQYLASRYDFIDLDRVGIHGSYGGYSLMALMORSDFRVAIAGA 763

QY      192  PVTLMIFPDGTGYTERYMGHPDNEOGYILGSVAMQAEKPPSPNNLLLLHGFLENVHFA 251
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      764  PVTLMIFPDGTGYTERYMGHPDNEOGYILGSVAMQAEKPPSPNNLLLLHGFLENVHFA 823

QY      252  HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYLOENLGSRIALAKYI 310
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      824  HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYLOENLGSRIALAKYI 882
```

```
RESULT 5
US-09-976-674-3
/ Sequence 3, Application US/09976674
```

```
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Qi, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Junten, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 863
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-976-674-3
```

```
Query Match      65.1%; Score 1093.5; DB 10; Length 863;
Best Local Similarity 57.3%; Pred. No. 5.3e-104;
Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY      1  FEGTDSPLEHLLVYVSVNPEGEVRLTDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      515  FEGTDSPLEHLLVYVSVNPEGEVRLTDRGYSHSCCISQHCDFRISKSNQKNPHCVSL 574

QY      61  YLSSPEDDPCTCKTEFNATLDSAGPLPDYTPPEIFSFESTTGTGMLTKPHDLOG 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      575  YLSSPEDDPCTCKTEFNATLDSAGPLPDYTPPEIFSFESTTGTGMLTKPHDLOG 634

QY      121  KKYPTLVFIYGGPQ----- 134
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      635  KKYPTLVFIYGGPQVQLVNNRFKGVKVFRLNTLASLGYVVVVIDNRGSCHRGLKEGAFK 694

QY      135  ---GOIEIDDOVEGIQYLASRYDFIDLDRVGIHGSYGGYSLMALMORSDFRVAIAGA 191
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      695  NQMGVEIEDDOVEGIQYLASRYDFIDLDRVGIHGSYGGYSLMALMORSDFRVAIAGA 754

QY      192  PVTLMIFPDGTGYTERYMGHPDNEOGYILGSVAMQAEKPPSPNNLLLLHGFLENVHFA 251
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      755  PVTLMIFPDGTGYTERYMGHPDNEOGYILGSVAMQAEKPPSPNNLLLLHGFLENVHFA 814

QY      252  HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYLOENL 300
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      815  HTSILSLFLVRAKGPYDLQIYPOERHSIRVPESGGEHYELHLHLAYLOENL 863
```

```
RESULT 6
US-09-976-674-23
/ Sequence 23, Application US/09976674
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Qi, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Junten, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 23
/ LENGTH: 892
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-976-674-23
```

```
Query Match      65.1%; Score 1093.5; DB 10; Length 892;
```

Best Local Similarity 57.3%; Pred. No. 5,6e-104;
Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

```

QY 1 FEGTDSPLEHLLVYVSYNPEGEVRLTDRGSHSCCISQHCDFPISXSNQKNPHCVSL 60
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 544 FQGTDTPTLEHLLVYVSYNAGEIVRLTTPGFSSCSMSQNDMFVSHSVSTPCVHV 603
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 YLSSPEDPTCKTEFMATILDSAGPLDPYTPPEIFSFESTGTGFLYGMCLKPPHLOPG 120
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 604 YLSSGEDDDPLHKORPFASMEASCPDPYVPELFFHHTSDVRLYGMIKKPHALOPG 663
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 KKYPTVLFLYGGPQ----- 134
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 664 KKHPTVLFLYGGPQVOLVNNSEFKIKYLRNLTLASLGVAVVVIDRGSCQGRGLRREGALK 723
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGHGMSYGGYLSLMALMORSDFRVALIGA 191
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 724 NQMGQVEIEDOVEGLOFVAEKYGFIDLSRVAIHGMSYGGFLSMGLIHKRPQVKVAILIGA 783
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 192 PVTLMIFDYTGTERYMGHPDQNEQGYLGSVAMQAEKPSBPENRLLILHGFLENVHFA 251
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 784 PVTVMAYDTGTERYMDVPENNGHYEAGSVLAHVEKLPNEPNRLLILHGFLENVHFF 843
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVPSGGEHYELHLHYQENL 300
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 844 HTNPLVSQILIRAGKPYOLOIYPNERHSIRCPESGEHYEVTLLHFLQEYL 892
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

RESULT 7
US-09-976-674-27

; Sequence 27, Application US/09976674
; Patent No. US20020115843A1

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve

; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre

; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV

; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674

; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27

; LENGTH: 892
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-976-674-27

Query Match 65.1%; Score 1093.5; DB 10; Length 892;
Best Local Similarity 57.3%; Pred. No. 5,6e-104;
Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

```

QY 1 FEGTDSPLEHLLVYVSYNPEGEVRLTDRGSHSCCISQHCDFPISXSNQKNPHCVSL 60
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 544 FQGTDTPTLEHLLVYVSYNAGEIVRLTTPGFSSCSMSQNDMFVSHSVSTPCVHV 603
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 61 YLSSPEDPTCKTEFMATILDSAGPLDPYTPPEIFSFESTGTGFLYGMCLKPPHLOPG 120
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 604 YLSSGEDDDPLHKORPFASMEASCPDPYVPELFFHHTSDVRLYGMIKKPHALOPG 663
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 121 KKYPTVLFLYGGPQ----- 134
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 664 KKHPTVLFLYGGPQVOLVNNSEFKIKYLRNLTLASLGVAVVVIDRGSCQGRGLRREGALK 723
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGHGMSYGGYLSLMALMORSDFRVALIGA 191
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 724 NQMGQVEIEDOVEGLOFVAEKYGFIDLSRVAIHGMSYGGFLSMGLIHKRPQVKVAILIGA 783
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 192 PVTLMIFDYTGTERYMGHPDQNEQGYLGSVAMQAEKPSBPENRLLILHGFLENVHFA 251
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

Db 784 PVTVMAYDTGTERYMDVPENNGHYEAGSVLAHVEKLPNEPNRLLILHGFLENVHFF 843
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVPSGGEHYELHLHYQENL 300
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 844 HTNPLVSQILIRAGKPYOLOIYPNERHSIRCPESGEHYEVTLLHFLQEYL 892
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

RESULT 8
US-09-976-674-33

; Sequence 33, Application US/09976674
; Patent No. US20020115843A1

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve

; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre

; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV

; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674

; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33

; LENGTH: 879
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-976-674-33

Query Match 64.7%; Score 1086.5; DB 10; Length 879;
Best Local Similarity 57.2%; Pred. No. 2,9e-103;
Matches 199; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

```

QY 2 EGTKDSPLEHLLVYVSYNPEGEVRLTDRGSHSCCISQHCDFPISXSNQKNPHCVSLX 61
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 532 KGTDTPTLEHLLVYVSYNAGEIVRLTTPGFSSCSMSQNDMFVSHSVSTPCVHV 591
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 62 YLSSPEDPTCKTEFMATILDSAGPLDPYTPPEIFSFESTGTGFLYGMCLKPPHLOPGK 121
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 592 KLSGEDDDPLHKORPFASMEASCPDPYVPELFFHHTSDVRLYGMIKKPHALOPGK 651
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 122 KKYPTVLFLYGGPQ----- 134
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 652 KKHPTVLFLYGGPQVOLVNNSEFKIKYLRNLTLASLGVAVVVIDRGSCQGRGLRREGALK 711
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGHGMSYGGYLSLMALMORSDFRVALIAGP 192
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 712 QMGQVEIEDOVEGLOFVAEKYGFIDLSRVAIHGMSYGGFLSMGLIHKRPQVKVAILIAGP 771
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 193 VTLMIFYDTGTERYMGHPDQNEQGYLGSVAMQAEKPSBPENRLLILHGFLENVHFAH 252
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 772 VTVVMAYDTGTERYMDVPENNGHYEAGSVLAHVEKLPNEPNRLLILHGFLENVHFFH 831
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 253 TSILSFLVRAGKPYDLOIYPOERHSIRVPSGGEHYELHLHYQENL 300
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 832 TNPLVSQILIRAGKPYOLOIYPNERHSIRCPESGEHYEVTLLHFLQEYL 879
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

RESULT 9

US-09-976-674-35

; Sequence 35, Application US/09976674
; Patent No. US20020115843A1

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve

; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre

; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV

; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674

; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117

;; PRIOR FILING DATE: 2000-10-12
;; NUMBER OF SEQ ID NOS: 61
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 35
;; LENGTH: 879
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-976-674-35

Query Match 64.7%; Score 1086.5; DB 10; Length 879;
Best Local Similarity 57.2%; Pred. No. 2.9e-103;
Matches 199; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 2 EGTGDSPLFHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 61
DB 532 KOTKOTPLEHHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 591
QY 62 KLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLQPG 121
DB 592 KLSGDDDDPLHKQPRFMSAMEAASCPDPYVPPIFHFTRSDVRLYGMVYKPHALQPG 651
QY 122 KYPPTVLFYIGGPQ----- 134
DB 652 KHPPTLVFYGGPQV 711
QY 135 --GQIEIDQVYEGLOYLASRYDFIDLDVGVIGMSYGYLSLMALMORSDFRVAIAGAP 192
DB 712 OMGOVEIEDQVYEGLOFVAEKYGFIDLSRVAIHGMSYGGFSLMGLIHKQVFKVAIAGAP 771
QY 193 VTLWTFYDTGYTERYMGHPDQNEOGYVYLSVAMQAEKFPSEPNRLLLHGFLENVHFAH 252
DB 772 VTVWMAVDTGYTERYMGHPDQNEOGYVYLSVAMQAEKFPSEPNRLLLHGFLENVHFAH 831
QY 253 TSILSLFVVRAGKPRPDLOIYPERHSIRPESGEHYELHLHYLOENL 300
DB 832 TNFLVSQILIRAGKPYQLOIYPERHSIRPESGEHYELHLHYLOENL 879

RESULT 10
US-10-237-271-3
;; Sequence 3, Application US/10237271
;; Publication No. US20030096328A1
;; GENERAL INFORMATION:
;; APPLICANT: THE BURNHAM INSTITUTE
;; APPLICANT: SMITH, Jeffrey W.
;; APPLICANT: KRIDEL, Steven J.
;; APPLICANT: AXELROD, Funko T.
;; TITLE OF INVENTION: SERINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS
;; FILE REFERENCE: BURN1100-1
;; CURRENT APPLICATION NUMBER: US/10/237,271
;; PRIOR FILING DATE: 2002-09-04
;; PRIOR APPLICATION NUMBER: US 60/317,842
;; PRIOR FILING DATE: 2001-09-06
;; NUMBER OF SEQ ID NOS: 19
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 3
;; LENGTH: 508
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-237-271-3

Query Match 61.5%; Score 1032.5; DB 15; Length 508;
Best Local Similarity 54.2%; Pred. No. 5.1e-98;
Matches 194; Conservative 43; Mismatches 54; Indels 67; Gaps 3;
QY 1 FEETKDSPLFHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 60
DB 160 FOGTKDPLFHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 219
QY 61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLQPG 120
DB 220 YKLSGDDDDPLHKQPRFMSAMEAASCPDPYVPPIFHFTRSDVRLYGMVYKPHALQPG 270

QY 121 KYPPTVLFYIGGPQ----- 134
DB 271 KHPPTLVFYGGPQV 330
QY 135 --GQIEIDQVYEGLOYLASRYDFIDLDVGVIGMSYGYLSLMALMORSDFRVAIAGAP 191
DB 331 NMGQVEIEDQVYEGLOFVAEKYGFIDLSRVAIHGMSYGGFSLMGLIHKQVFKVAIAGAP 390
QY 192 VTLWTFYDTGYTERYMGHPDQNEOGYVYLSVAMQAEKFPSEPNRLLLHGFLENVHFAH 251
DB 391 VTVWMAVDTGYTERYMGHPDQNEOGYVYLSVAMQAEKFPSEPNRLLLHGFLENVHFAH 450
QY 252 TSILSLFVVRAGKPRPDLOIYPERHSIRPESGEHYELHLHYLOENL 300
DB 451 TNFLVSQILIRAGKPYQLOIYPERHSIRPESGEHYELHLHYLOENL 508

RESULT 11
US-09-976-674-19
;; Sequence 19, Application US/09976674
;; Patent No. US20020115843A1
;; GENERAL INFORMATION:
;; APPLICANT: Qi, Steve
;; APPLICANT: Akinsanya, Karen
;; APPLICANT: Riviere, Pierre
;; APPLICANT: Julien, Jean-Louis
;; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
;; FILE REFERENCE: 70669
;; CURRENT APPLICATION NUMBER: US/09/976,674
;; CURRENT FILING DATE: 2001-10-12
;; PRIOR APPLICATION NUMBER: US 60/240,117
;; PRIOR FILING DATE: 2000-10-12
;; NUMBER OF SEQ ID NOS: 61
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 19
;; LENGTH: 658
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-976-674-19

Query Match 44.0%; Score 739; DB 10; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.8e-67;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FEETKDSPLFHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 60
DB 524 FEETKDSPLFHLHYVSYVNPGEVTRLTDRGYSHCISQHCDFISKYSNOKNPHCVSL 583
QY 61 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLQPG 120
DB 584 YKLSSEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTGTFTLYGMLYKPHDLQPG 643
QY 121 KYPPTVLFYIGG 132
DB 644 KYPPTVLFYIGG 655

RESULT 12
US-09-976-674-11
;; Sequence 11, Application US/09976674
;; Patent No. US20020115843A1
;; GENERAL INFORMATION:
;; APPLICANT: Qi, Steve
;; APPLICANT: Akinsanya, Karen
;; APPLICANT: Riviere, Pierre
;; APPLICANT: Julien, Jean-Louis
;; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
;; FILE REFERENCE: 70669
;; CURRENT APPLICATION NUMBER: US/09/976,674
;; CURRENT FILING DATE: 2001-10-12
;; PRIOR APPLICATION NUMBER: US 60/240,117
;; PRIOR FILING DATE: 2000-10-12
;; NUMBER OF SEQ ID NOS: 61

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-976-674-11

```

Query Match	44.0%;	Score 739;	DB 10;	Length 661;
Best Local Similarity	100.0%;	Pred. No. 1.8e-67;		
Matches 132;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	FECTKDSPLEHNHLYVSVYNPGCVRLTLDRGYSHSCCISQCHDFISXYSQKMPHCYSL	60
Db	524	FECTKDSPLEHNHLYVSVYNPGCVRLTLDRGYSHSCCISQCHDFISXYSQKMPHCYSL	583
Qy	61	YKLSSEDDPTCKTKRFMAITLDSAGPLADYPRPIFSFESTGTGLYGMLYKHNDLQPG	120
Db	584	YKLSSEDDPTCKTKRFMAITLDSAGPLADYPRPIFSFESTGTGLYGMLYKHNDLQPG	643
Qy	121	KKPTVLFIYGG	132
Db	644	KKPTVLFIYGG	655

```

RESULT 13
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;
; APPLICANT: O'i Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Julien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
;
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;
; LENGTH: 690
;
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-976-674-7

```

Query Match	44.0%;	Score 739;	DB 10;	Length 690;
Best local Similarity	100.0%;	Pred. 719;	DB 10;	Length 690;
Matches 132;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	FECTKDSPLEHLLYYVSVYNPGEVTRLTDRCYSHSCCISQHCDFEISKYSNQNPCHVSL	60
Db	524	FECTKDSPLEHLLYYVSVYNPGEVTRLTDRCYSHSCCISQHCDFEISKYSNQNPCHVSL	583
QY	61	YKLSPEDDPFCCKTEFMATIIIDSGPLPDYTPPEISFBSSTGFTLYGMLYKPHLDQPS	120
Db	584	YKLSPEDDPFCCKTEFMATIIIDSGPLPDYTPPEISFBSSTGFTLYGMLYKPHLDQPS	643
QY	121	KKYPVTLFTYGG	132
Db	644	KKYPVTLFTYGG	655

RESULT 14
 US-09-976-674-29
 ; Sequence 29, Application US/09976674
 ; Patent No. US20020115843A1
 GENERAL INFORMATION:
 APPLICANT: Qi, Steve
 APPLICANT: Akinsanya, Karen
 APPLICANT: Riviere, Pierre
 APPLICANT: Nunien, Jean-Louis

```

: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
:
: FILE REFERENCE: 70669
:
: CURRENT APPLICATION NUMBER: US/09/976,674
:
: CURRENT FILING DATE: 2001-10-12
:
: PRIOR APPLICATION NUMBER: US 60/240,117
:
: PRIOR FILING DATE: 2000-10-12
:
: NUMBER OF SEQ ID NOS: 61
:
: SOFTWARE: Patentin version 3.1
:
: SEQ ID NO 29
:
: LENGTH: 832
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
US-09-976-674-29

```

	Query Match	37.5%	Score 629.5;	DB 10;	Length 832;
	Best Local Similarity	48.0%;	Pred. No. 5.4e-56;		
	Matches 117;	Conservative 30;	Mismatches 48;	Indels 49;	Gaps 1
Qy	1	FEGTADSPLENNHLYVSVYNPGEVTRLTDRGYSHCISCCIQHCDFFISKYNSNQKNPHCVSL	60		
Db	544	FOITKDTPLNHLHYVSYEAAGEIVRLTPPGFSCHSCMSONFDMFVSHVSSTPRPCNVH	603		
Qy	61	YKLSSPEDPTCKTEPFATLIDSGRLPDVTPEIEFSEFTGTGLYGMILYRPHDLPQG	120		
Db	604	YTLGGDDDLHKQRFPMASMEAAACRPDVYPPEIHFHTRSRVLRKYGIYKFRHALQFG	663		
Qy	121	KKPYLVLFYGGPO-----	134		
Db	664	KKHPPVLFPYGGPQQQLVNNSFKGIKKYLRLNTLASLGVAVVVIDGRGSCORGRLFEQALK	723		
Qy	135	---GOEIIDDGVEGLQYLASRYDFIDLDRVIGHMSSYGYSLSMALMQRSDIFRWAIAGA	191		
Db	724	NQMGGVEIEDQGVLEGQLFAEKYGFIDLSRVAIHMSSVGFLSLMGILHKQVFKAPRLAV	783		
Qy	192	PVTL 195			
Db	784	PPRL 787			

```

RESULT 15
US-09-976-674-31
; Sequence 31, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Ol, Steve
; APPLICANT: Akhsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Unien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
; LENGTH: 832
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-976-674-31

```

[illegible]

QY	121	KKYPTVLFYGGPQ-----	134
Db	664	KKHPTVLFYVGGPQVOVUNNSFKGICKYLRNLTLASLGAVVWIDGRGSCQRGILRFBGALK	723
QY	135	---GOIEIDDOVEGLQYLASRYDIDLDRVGIHGWMSYGYLSLMALMORSDFRVAIAGA	191
Db	724	NOMGOVEIEDQVEGLQFVAEKYGFIDLSRVVAIHGWMSYGFSLMGLIHKPQVFKQPLAY	783
QY	192	PVTL	195
Db	784	PPRL	787

Search completed: October 15, 2003, 17:28:49
Job time : 44.8027 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 8.14576 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-3
Perfect score: 1680

Sequence: 1 FEGTKDSPLEHLLVYVSVN.....HLHLTYQENLGSRIALKVI 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	502	29.9	931	2 T32919	hypothetical prote
2	484.5	28.8	738	2 A87516	dipeptidyl peptida
3	398.5	23.7	741	2 UC5142	X-Pro dipeptidyl-p
4	379	22.6	711	2 S66261	X-Pro dipeptidyl-p
5	359.5	21.4	793	2 T41703	dipeptidyl aminope
6	342.5	20.4	760	1 S23752	dipeptidyl-peptida
7	340.5	20.3	792	1 A39914	dipeptidyl-peptida
8	337.5	20.1	766	1 CDH026	dipeptidyl-peptida
9	314	18.7	931	2 A49737	dipeptidyl aminope
10	310	18.5	818	1 A30107	dipeptidyl aminope
11	303.5	18.1	779	2 T25173	hypothetical prote
12	303.5	18.1	799	2 T25174	hypothetical prote
13	284.5	16.9	829	2 T19514	hypothetical prote
14	262	15.6	803	2 A41793	dipeptidyl aminope
15	248	14.8	803	2 T68600	dipeptidyl aminope
16	248	14.8	865	2 T54331	dipeptidyl aminope
17	245.5	14.6	793	2 F82858	dipeptidyl-peptida
18	244	14.5	743	2 T37700	probable dipeptidyl
19	226	13.5	631	2 H75007	probable acylamino
20	209.5	12.5	683	2 E87495	prolyl oligopeptid
21	207	12.3	642	2 C71137	hypothetical prote
22	188	11.2	709	2 B82580	alanyl dipeptidyl
23	179.5	10.7	759	2 T38593	fibroblast activat
24	171	10.2	591	2 H72474	probable acylamino
25	168	10.0	622	2 F71174	hypothetical prote
26	167	9.9	632	2 E70057	peptidase PA01418
27	152.5	9.1	657	2 E70025	probable acylamino
28	151.5	9.0	569	2 S74053	probable acylamino
29	149	8.9	536	2 F90299	acylaminoacyl-pept

30	147	8.8	721	2 T09631	probable acylamino
31	146	8.7	608	2 F83397	hypothetical prote
32	143	8.5	828	2 G87584	hypothetical prote
33	140.5	8.4	572	2 F72455	probable acylamino
34	140	8.3	732	1 JC4655	acylaminoacyl-pept
35	138	8.2	732	1 S07624	acylaminoacyl-pept
36	137	8.2	265	2 B84063	hypothetical prote
37	136.5	8.1	667	2 A87711	prolyl oligopeptid
38	135.5	8.1	598	2 F84199	hypothetical prote
39	127	7.6	618	2 D87651	probable acylamino
40	126.5	7.5	582	2 D72636	acylaminoacyl-pept
41	126.5	7.5	732	1 JU0132	esterase (imported
42	129.5	7.4	4	2 AG1816	conserved hypotet
43	121.5	7.2	891	2 G82543	acylaminoacyl-pept
44	116	6.9	674	2 B84381	prolyl oligopeptid
45	115	6.8	642	2 D87379	

ALIGNMENTS

RESULT 1					
T32919					
hypothetical protein K02F2.1 - Caenorhabditis elegans					
C/Species: Caenorhabditis elegans					
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #ext_change 29-Oct-1999					
C/Accession: T32919					
R/Magdi, L.; Goela, D.					
Submitted to the EMBL Data Library, January 1998					
A/Description: The sequence of C. elegans cosmid K02F2.					
A/Reference number: Z21246					
A/Accession: T32919					
A/Status: preliminary; translated from GB/EMBL/DBJ					
A/Molecule type: DNA					
A/Residues: 1-931 <MAG>					
A/Cross-references: EMBL:AF043699; PIDN:AB97564.1; GSPDB:GN00019; CESP:K02F2.1					
A/Experimental source: strain Bristol N2; clone K02F2					
A/Genetic: CESP:K02F2.1					
A/Map position: 1					
A/Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2; 5					
Query Match					
Best local similarity 33.2%; Score 502; DB 2; Length 931;					
Matches 122; Conservative 56; Mismatches 104; Indels 86; Gaps 14;					
QY	1	FEGTKDSPLEHLLVYVSVN	PGEVTRLTD	RGYSHSCISQHCDFISKYSNOK	---NPH 56
DB		563 YVANSHPLENNI-CVSHYRTQGHQLTESGI	-----C--FKSERANGKALDLDH	610	
QY	97	CVSLYKLS--SPEDDPTCKTEP-W	-----ATILDSAGP-LPD--YTPPEIFS	98	
DB		611 GPACVMTSVGSPAE--CRFVSFRWKENELVSTVYANITVSGHGGQDLHFDSPEMIE	667		
QY	99	FES-TTGFLLVQMLKPPHLDOPKCYPTVLFTYGGP	-----	133	
DB		668 FQSKKTGLMHTVAMILPSPNFDYKRYPVFHVYVGGGIQIVANDFSMIQYIRFCRLGYVV	727		
QY	134	-----OQOIEIDDOVEGLOYLASRY-DPFLDRLVGIHGSYCG	170		
DB		728 VFDNRGSAHRCIEFERHIIKKMGIVVEDVQEGLOMLAERKGFPMDSRVVHHGMSYCG	787		
QY	171	YLSLMAIORSIDIFVVALAGAPVTIMFTDGYTERYWGHPDNOGGYLGSVANQAEKF	230		
DB		788 YMALQMIKHPNIVPAALAGAVSDMRVYDTVYTERYMGYP-LSEHVVGASSITGLVEXL	846		
QY	231	FSEPRLLHLHGFLDENVFAHTSITLSFLVAGKPYDLOIYPOEHSIRVSESGHYVL	290		
DB		847 PDEPRRLVHLHGMDENVHFAHLTLHVDCEIKKGKHELVIFPNERHGVRRNDASITYLDA	906		
QY	291	HLHLTYQOE 298			
DB		907 RMYFYAQO 914			

RESULT 2

AB7516
 dipeptidyl peptidase IV [Imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: AB7516
 R.Nietman, W.C.; Fejldjy, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heideberg, J.B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of Caulobacter crescentus.
 A/Reference number: AB7249; MUID:21173698; PMID:11259647
 A/Accession: AB7516
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-738 <STO>
 A/Cross-references: GB:AE005673; NID:g13423647; PIDN:AAK24125.1; GSPDB:GND0148
 C/Genetics:
 A/Gene: CC2154

Query Match 28.8%; Score 484.5; DB 2; Length 738;
 Best Local Similarity 32.2%; Pred. No. 5e-34;
 Matches 106; Conservative 53; Mismatches 109; Indels 61; Gaps 6;
 1 FEETKDSPLEHNLVYVSVNPEVTRLTDRGYSHSCISQHCDFITSKYSNQKNPCHVSL 60
 394 FMSITPIERRLYEYSYAKFKPKALTSAGGMMAKVADNGAFACTYSDPTPSOTAL 453
 61 YKLSSP-----EDPTCKTEFWATILDSAGPLPDYTPPEIFSESTTGTLYGMLYKPH 115
 454 YSADGKRWIRIENKLAEGHPW---PYANL PQ--PFGSLKADGETLHYELKPI 506
 116 DLQPGKKYPTVLTLYGPGQ----- 134
 507 GFDPACKYPAIVSVYGGPHQORVMKNHSPSEERTYEAGYVIFKLDNRGSGNSAFEMRA 566
 135 -----COIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYSLMALMQRSDIFRVAIA 189
 567 LDRKLGIVEVEDLGLAKFLASQ-PYVDADKLGVMGSYSGFMALMLTMENTPFFAGAA 625
 190 GAVPTLWIFYDTGTERYMGHPDQNEGYVLGSVAMQAEKFPSEPNRLLLHGFLEDNVH 249
 626 GADPTESLYDTAYTERYMGKPDENKAGYVSDINNRIDL--APGSLLLHGMADDNVI 683
 250 FAHTSILSLFLVRAKGYDQIYPOERHS 278
 684 FENSTRMLAALQKRAILFEMAMPGERHS 712

RESULT 3

JC5142
 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
 C/Species: Xanthomonas maltophilia
 C/Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
 C/Accession: JC5142
 R.Kabashima, T.; Ito, K.; Yoshimoto, T.
 J. Biochem. 120, 1111-1117, 1996
 A/Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expression
 A/Reference number: JC5142; MUID:97164011; PMID:9010758
 A/Accession: JC5142
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-741 <KAB>
 A/Cross-references: DDBJ:D83263; NID:g1753196; PIDN:BA11872.1; PID:g1753197
 C/Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl residues from peptides.
 C/Superfamily: dipeptidyl-peptidase IV
 C/Keywords: dipeptidyl-peptidase; hydrolase; membrane bound
 F/4-18/Domain: transmembrane #status predicted <TM>
 F/610/Active site: Ser #status predicted
 F/685/Active site: Asp #status predicted
 F/717/Active site: His #status predicted

Query Match 23.7%; Score 398.5; DB 2; Length 741;
 Best Local Similarity 28.0%; Pred. No. 1.5e-26;
 Matches 98; Conservative 57; Mismatches 120; Indels 75; Gaps 8;

1 FEETKDSPLEHNLVYVSVNPEVTRLTDRGYSHSCISQHCDFITSKYSNQKNPCHVSL 60
 396 FRAGRISARESGIYAVP-LQGGQPRQLSKAPKMSHSAFARNASVYVDSNNSNTPQIETL 454
 61 YKLSSPDDPTCKTEFWATIL--DSAGLPDY-----PPEISFESTTGT-LYGM 110
 455 FRANG-----EKITLVENDLADKHPHARYREARQRPVEFGTLTAADGKTPLNYS 504
 111 LYKPELQPGKKYPTVLTLYGPG----- 133
 505 VIKRPGFDPACKYPAIVSVYGGPASTVTYDMSWPRGDHLFNOYLAQGVVFLDNRGTP 564
 134 -----OQIEIDDOVEGLQYLASRYDFIDLRVGIHGSYGYSLMALMQR 180
 565 RRGDRFGALYKQGTVEVADQLRGVAML-KQGPWVDPARIGVGWNSNGYMTLMALAKA 623
 181 SDIFRVAIAGAPVTLWIFYDTGTERYMGHPDQNEGYVLGSVAMQAEKFPSEPNRLLL 240
 624 SDSYACGVAGAPVTDGLYDSHYTERYMDLPANDGVRDEARVLTIEGLRSP--LLLI 680
 241 HGFLEDNVHFAHTSILSLFLVRAKGYDQIYPOERHSIRVPSGCHYEL 290
 681 HGMADDNVLTFTNSTLSMALQKRGQPFELMTYFGAHGSLGADALHRYIV 730

RESULT 4

S66261
 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
 C/Species: Flavobacterium meningosepticum
 C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
 C/Accession: S66261
 R.Kabashima, T.; Ito, K.; Yoshimoto, T.
 Arch. Biochem. Biophys. 320, 123-128, 1995
 A/Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from F
 A/Reference number: S66261; MUID:95314307; PMID:7793970
 A/Accession: S66261
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-711 <KAB>
 A/Cross-references: EMBL:D42121; NID:g577283; PIDN:BAA07702.1; PID:g577284
 C/Superfamily: dipeptidyl-peptidase IV
 C/Keywords: dipeptidyl-peptidase hydrolase

Query Match 22.6%; Score 379; DB 2; Length 711;
 Best Local Similarity 26.6%; Pred. No. 7.2e-25;
 Matches 90; Conservative 58; Mismatches 118; Indels 72; Gaps 5;

19 VNPGEVTRLTDRGYSHSCISQHCDFITSKYSNQKNPCHVSLYKLSSP-----DDPT 71
 390 INTGKQLSLNMGNNNSAASFKNFNFINTSTAKAPTYTLKDAAGKVKELQNNDDL 449
 72 CKTKFEMATILDSAGPLDYTPPEIFSESTTGTLYGMLYKPHDQPGKKYPTVLFYIG 131
 450 NKLKS-----DNFIAKEFTITMAADQNNAMIKRPNDRPAKTYVVFNFQYS 497
 132 GP-----OQIEIDDO 142
 498 GPGSQGVANSWDGNGIWFMDLAQKGYLVAVCDGRGTGRGTQKTKYTKYLKSLKEIDEI 557
 143 VEGLOYLASRYDIDLRVGIHGSYGYSLMALMQRSDIFRVAIAGAPVTLWIFYDQ 202
 558 ITAAKRLGQO-STVDSRIGIFGMSYGGFMASLAKGADVFMGIAVAPVTNMRFTDST 616
 203 YTERYMGHPDQNEGYVLGSVAMQAEKFPSEPNRLLLHGFLEDNVHFAHTSILSLFLV 262
 617 YTERFLQGTQENKDDGDVNSPTTYAKLL--KGKFLIHGTADDNVHFNQNSMEFSALIQ 673
 263 AKGPYDQIYPOERHSIRVPSGCHYELHLHYLOENT 300

Db 674 NKGQDFMAYPDKNHSITIGNTRPOLYKEMTNVLENTL 711

RESULT 5

T41703

dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

T41703

R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1998

A:Accession: T41703

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-793 <NTR>

A:Cross-references: EMBL:AL031180; P1DN:CAA20138.1; GSPDB:GNO0066

A:Experimental source: strain 972h-; cosmid c2B11 -chimeric

C:Genetics:

A:Gene: SPAC2B11.08

A:Map position: 1

C:Superfamily: dipeptidyl-peptidase IV

Query Match 21.4%; Score 359.5; DB 2; Length 793;
Best Local Similarity 31.5%; Pred. No. 4,1e-23;
Matches 113; Conservative 43; Mismatches 128; Indels 75; Gaps 17;

5 KDSPLHLLVSVYVNGEYTRLTDRGYS--HSCCISQHCDFPISKYSNQKNPHCVSLY 61

442 KDS-TERHLYVVS-LDTLEIYGITDNGEDGYSTSPSPGDFVILVHGPDPV---WQ 495

62 KLSPPDDPTCKTEPFATILDSAGPLPDYTPPEI---FSFESTGFTLYGMLYKPHDL 117

496 ELKSTDKDQYCLSLSTNSRIKQ---LSITLPSVEYKGLTFNDT-L-FNF--MERRRNF 549

118 QPGKXPTVLYFIYGPQGG-----IEIDQVEG-----LQY 148

550 DVNKKTPVLFPAVGGPSQOVAKLFRVDFQAVLASHDPFEFIVTLDGRTGNGNAPRY 609

149 LMSR-----YD-----FTDLDRVGIGHGSGYGYSLMALMQRSDIFRYALA 189

610 SVSRHLGEMESYDQGAQKFMADLPFVDENHVGIMGWSYGYTLTKTL-ETQDVFSYGM 668

190 GAVTLMIFFDYDGYTERYMGHPDQNEGYLLGSVAMQAEKPESEPRLLHLHGFLENHA 249

669 VAPVTDMLRYDSYTERYMDLPQYNKEG-YKNSQIHDIYERF-KQLKRFVAHGTGDVNH 726

250 FAHTSILSLFLVRAG-KPYDLQIYPOERHSIRVPESGEHELHLHYLOENLGSRIAL 307

727 FQSHMLMGLNLANCVNDMAVFPDSASHI-----SYHNASLSIYRLSEWIDALGRI 781

RESULT 6

S23752

dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse

N:Alternate names: CD26 alpha subunit; THAM alpha subunit

C:Species: Mus musculus (house mouse)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S23752; A46465; A56030

R:Marquet, D.; Bernard, A.M.; Vliet, I.; Darmoul, D.; Naquet, P.; Pierres, M.

J. Biol. Chem. 267, 2200-2208, 1992

A:Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-di-

A:Reference number: S23752; MUID:92129288; PMID:1370813

A:Accession: S23752

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-760 <MAR>

A:Cross-references: EMBL:X58384

R:Vivier, I.; Marquet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;

J. Immunol. 147, 447-454, 1991

A:Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase

A:Reference number: A46465; MUID:91302787; PMID:1712807

A:Accession: A46465

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-20 <VIV>

A:Experimental source: M14.T thymoma cells, Swiss nu/nu

A>Note: Sequence extracted from NCBI backbone (NCBI:42236)

R:Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marquet, D.

Biochemistry 33, 15204-15214, 1994

A:Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.

A:Reference number: A56030; MUID:95092780; PMID:7999781

A:Accession: A56030

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 746-760 <BER>

A:Cross-references: GB:U12620

C:Genetics:

A:Gene: CD26

C:Superfamily: dipeptidyl-peptidase IV

C:Keywords: dipeptidylpeptidase hydrolase; glycoprotein; transmembrane protein

F:213,223,315,514,679/Binding site: carboxylate (Asn) (covalent) #status predicted

F:624,702,734/Active site: Ser, Asp, His #status predicted

Query Match 20.4%; Score 342.5; DB 1; Length 760;
Best Local Similarity 28.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 51; Mismatches 120; Indels 89; Gaps 14;

5 KDSPLHLLVSVYVNGEYTRLTDRGYS--HSCCISQHCDFPISKYSNQKNPHCVSLY 56

417 KEMPGRLPLYLHSTHDKELRVLE-----QLTD-HTVVKCLSDLPNERCQYAVSPSKAKYVQLG 465

57 C-----SVLYKSSPEDPTCKTEPFATILDSAG---PLPDYTPPE---IFSFESTGFT 106

466 CMGPGRLPLYLHSTHDKELRVLE-----DNSALDRMLQDVQMPKSLDFVILNETRF- 518

107 LYGMLYKPHDLQPGKXPTVLYFIYGP-----IEIDQVEG-----LQY 133

519 WYGMILPPLH-FDKSKKYPLLDVYAGPCGQKADASRLIMATYLASTENITVASFDGRGS 577

134 -----OCQIEIDQVEGLQYLAHYDFIDLRVGIHMSYGYSLMALMQ 179

578 GYGGD¹KIMHAINRLCTLEVEDQIEAARQV-¹KMGFVDSKRVAILMGWSYGYVTSMVLGS 636

180 RSDIFPVATAGAPVTLMIFFDYDGYTERYMG--HPDQNEGYLLGSVAMQAEKPESEPRNL 237

637 GSGVFCGIALVAVSWWEYDSYTERYMGRLPIPEDNLHYRNSYMSRAEHF--KOVEX 694

238 LLLHGFLENVFAHTSILSLFLVRAGKPYDLQIYPOERHSIRVPESGEHELHLHYLQ 297

695 LLHGTADNVNHFQGSASQISKVLVDAGVDFQAMWYTDDEHGIASSTAHQHIIYSHMSHFLQ 754

RESULT 7

A39914

dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat

N:Alternate names: GP110; membrane glycoprotein 110K; OX-61

N:Contains: dipeptidyl-peptidase IV soluble form

C:Species: Rattus norvegicus (Norway rat)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781

R:Hong, W.; Doyle, D.

Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987

A:Title: cDNA cloning for a bile canalculus domain-specific membrane glycoprotein of

A:Reference number: A39914; MUID:88068516; PMID:3479775

A:Accession: A39914

A:Molecule type: mRNA

A:Residues: 1-792 <HON>

A:Cross-references: GB:J02997; MUID:92044463; P1DN:AAA11272.1; P1D:G204464

R:Ogata, S.; Mizumoto, Y.; Ikehara, Y.

J. Biol. Chem. 264, 3596-3601, 1989

C:Genetics:
 A:Gene: GDB:DDP4
 A:Cross-references: GDB:125239; OMIM:102720
 A:Map position: 2q24.3-2q24.33
 A:Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; proteinase; transmembr
 F:1-6/Domain: intracellular #status predicted <INT>
 F:7-28/Domain: transmembrane #status predicted <TM>
 F:29-766/Domain: extracellular #status predicted <EXT>
 F:85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (asn) (covalent) #status
 F:630,708,740/Active site: Ser, Asp, His #status predicted

Query Match 20.1%; Score 337.5; DB 1; Length 766;
 Best Local Similarity 28.0%; Pred. No. 3.2e-21;
 Matches 99; Conservative 53; Mismatches 127; Indels 75; Gaps 13;

Qy 5 KDSPLEHLLVYVSVNPEVTRLTDRGYSHSCCIS-QHCDPFIKSYNQKNPH---C--- 57
 Db 423 KMPGGRNLVYKIQIDYTKVTCL-----SCELNPERQYVSFSEKAKYQLRCGSP 475

Qy 58 -VSLYKSSPEDDPCKTKKEFWATILDSAGLPDTPPE---TFSESTTGTLXGMLYK 113
 Db 476 GLPLVTLHSSVNDKGLRYLED-NSALDKM-LQNVQMPESKLDFTILNETKF-WYOMTLP 531

Qy 114 PHDLQPGKKYPTVLFYGGP----- 133
 Db 532 PH-FPKSKKYPLLLVNVAGPSCQKADYVRLMWTATYSTENITVASPDGSGSGYQGDKI 590

Qy 134 -----OQIEIDDOVEGLQYLAIRYDFIDLRVGIHWSYGYLSLWALMQRSDFRV 186
 Db 591 MHAIRRLGTPEVEDEQIEAARQF-SKMGFVNDKRIALIMQMSGVYVTWVLGSSGSGVKC 649

Qy 187 AIAAGPVTLMFYDGTGYERYKMG--HPDNEGGYLLGSYANQAEKFPSEPNRLILHGL 244
 Db 650 GIAVAPVSRMEYDSEVYERYWGLPTPEBDNLHYNSSTYMSRAENF--KOVEYLLIHGTA 707

Qy 245 DENVFATSTILSLFLVRAKPYDQIYQERHSIRVPESEGEYELHLHYLOE 298
 Db 708 DDNVHFGQSAQISKALVDGVDFQAMWYTDDEDHGIASSTAHQHITTHMSHFIKQ 761

RESULT 9
 A49737
 dipeptidyl aminopeptidase (EC 3.4.14.-) - Yeast (Saccharomyces cerevisiae)
 N:Alternate names: protease O5045; protein YOR196c; protein YOR50-9
 C:Species: Saccharomyces cerevisiae
 C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 29-Oct-1999
 C:Accession: A49737; S45451; S60946; S67112; S71721
 R:Santa Anna-A. S.; Herskowitz, I.
 submitted to the Protein Sequence Database, July 1993
 A:Reference number: A49737
 A:Accession: A49737
 A:Molecule type: DNA
 A:Residues: 1-931 <SN>
 A:Cross-references: GB:L21944; NID:G347196; PIDN:AAA5119.1; PID:G347197
 R:Anna-Artiola, S.S.; Herskowitz, I.
 Yeast 10, 801-810, 1994
 A:Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase
 A:Reference number: S45451; MUID:95066382; PMID:7975897
 A:Accession: S45451
 A:Molecule type: DNA
 A:Residues: 1-931 <ANN>
 A:Cross-references: EMBL:L21944; NID:G347196; PIDN:AAA5119.1; PID:G347197
 R:Galison, F.; Dujon, B.
 submitted to the EMBL Data Library, October 1995
 A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome
 A:Reference number: S60938
 A:Accession: S60938
 A:Molecule type: DNA
 A:Residues: 1-931 <GAL>
 A:Cross-references: EMBL:X92441; NID:G1050762; PIDN:CAA63182.1; PID:G1050771
 R:Boyer, J.; Fairhead, C.; Gallion, L.; Galison, F.; Michaux, G.; Thierry, A.; Dujon, B

submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67104
 A:Accession: S67112
 A:Molecule type: DNA
 A:Residues: 1-931 <BOY>
 A:Cross-references: EMBL:275127; NID:G1420507; PIDN:CAA99437.1; PID:E252398; PID:G14205
 A:Experimental source: strain S288C
 R:Galison, F.; Dujon, B.
 Yeast 12, 877-885, 1996
 A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
 A:Reference number: S71721; MUID:96437977; PMID:8840505
 A:Accession: S71721
 A:Molecule type: DNA
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Cross-references: SGD:S0005745; MIPS:YOR219C
 A:Map position: 15R
 A:Gene: SGD:STE13; YC11
 A:Cross-references: EMBL:X92441; NID:G1050762; PIDN:CAA63182.1; PID:G1050771
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
 C:Genetics:
 A:Function: involved in processing of alpha-factor prepropheromone
 A:Description: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
 F:113-150/Domain: transmembrane #status predicted <TM>
 F:377/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:785,863,896/Active site: Ser, Asp, His #status predicted

Query Match 18.7%; Score 314; DB 2; Length 931;
 Best Local Similarity 25.6%; Pred. No. 4.6e-19;
 Matches 100; Conservative 54; Mismatches 133; Indels 104; Gaps 12;

Qy 1 PEGTKDSPLEHLLVYVSV-----NPGVTRLTDRGYSHSCCISQHCDFISK 48
 Db 562 FTANEIGVMSOHLVYSISLTDSTQNTFQSLQNP-----SKYDFYDELSSARVATSK 615

Qy 49 YSNQKNP-----HCVSLYKSSPEDDPCKTKKEFWATILDSAGLPDYT 92
 Db 616 KLGPPTPIKAGAPLTRVNLVAEIHDDSLQLTKDE---KFE-----KIKVND 660

Qy 93 PPEIFSEST---GFTLYGLYKPHDLQPGKKYPTVLFYGGPQO----- 136
 Db 661 LP-ITSYKTVWLDQVEEINYEIKPANLPKKKPIVLNVIYGPSQOFTTKSSLAPEQA 719

Qy 137 -----IIDDQVGS-----LOYLASRYDFIDLRV 161
 Db 720 VVSGLDVIVLQIEPRGTGKMSFRSWAREKLGWEPDITEVTKFQRSQHIDESKI 779

Qy 162 GIHGWSYGYSLWAL-MQRSDFRVALAGAPVTLWIFYDGTGYERYWGHDPQNGQYLL 220
 Db 780 AIMGWSYGYFTSLKTVELNDGDTFRVAVAPVNTWTLVDSVYERYWYNQPSSEHGEYFE 839

Qy 221 GSVAMQAEKFPSEPNRLILHGLFDENVHFAHTSILSLFLVRAG-KPYDQIYQERHSI 279
 Db 840 VSTIQNFKSPES-LKRLFIYVGTFTDDNVHIGNTRLVQGLMLGLTNYDMHIFPDSHSI 898

Qy 280 RVPESEGEYELHLHYLOENIGSRIALAKVT 310
 Db 899 RYHNAORIVFQKLYVWLRDAPAFERFDTEVL 929

RESULT 10
 A30107
 dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protease YHR028c
 C:Species: Saccharomyces cerevisiae
 C:Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
 C:Accession: S46780; A30107
 R:Du, Z.
 submitted to the EMBL Data Library, June 1994
 A:Description: The sequence of S. cerevisiae cosmid 8082.
 A:Reference number: S46773
 A:Accession: S46780

A:Molecule type: DNA
 A:Residues: 1-818 <DUB>
 A:Cross-references: EMBL:U10399; NID:9500689; PIDN:AA68879.1; PID:9500698; GSPDB:GN0000
 R.Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.
 J. Cell Biol. 108, 1363-1373, 1989
 A>Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an in
 A:Reference number: A30107; MUID:89174971; PMID:2647766
 A:Accession: A30107
 A:Molecule type: DNA
 A:Residues: 1-82, 'H', 84-124, 'N', 126-181, 'LRLET', 189-199, 'N', 201-365, 'DFRGRKRF', 376-57
 A:Cross-references: EMBL:X15484
 A:Note: the authors translated the codon ACC for residue 572 as Asn
 C:Genetics:
 A:Gene: SGD:DA2; STE13; MIPS:YHR028C
 A:Cross-references: SGD:S0001070; MIPS:YHR028C
 A:Map position: 8R
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptidase hydrolase; glycoprotein; transmembrane protein; yeast vacu
 F:30-45/Domin: transmembrane #status predicted <TM>
 F:63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.5%; Score 310; DB 1; Length 818;

Best Local Similarity 26.9%; Pred. No. 8.7e-19; Indels 106; Gaps 13;
 Matches 98; Conservative 38; Mismatches 122;

QY 1 PEGTDSPLEHLLVYVSVNPGVTRLTDRGYS--HSCCISQHCDFISKYSNOKNPH-- 56
 DB 451 FISTKSSSTERYVYVYIDLASPNBIIIVTDTSDEGVYDVFSGRRGLTYGPKVPYQK 510
 QY 57 -----C-----VSLYKLSPPEDDPCKTKERWATITLDSAGLPTPTPEIRIS 98
 DB 511 IVDHRSKAKCKDKGNVLAKSLYHLEKNE-----VLTKI--LEDYAVPRKSP 555
 QY 99 FESTGTFTLYG--MLYKPHDLP-----GKKYPTVLYGAPQ----- 134
 DB 556 RELNLCKDFGDIIVNSVEILPNDEFTLSDIHYPIFFRAQSPNQVVKTFSGVCFNEV 615
 QY 135 -----GQIEIDDOVEGLQYLASRYFDIDRVG 162
 DB 616 VASQNALIVVVDGRGTGFGQDFRSLVDRLDYFARQIISAALYGS-LTFVDPQKIS 674
 QY 163 IHGWSYGYLSLMAIMQGRD-I-FRVALIAGAPVTLMEFYDTGTERWGMHPDQEGCYLYG 221
 DB 675 LFGWSYGYLTLTLEKDGGRHKYGMSPVPTDMKVFYVTERYMTPOENFGDYVES 734
 QY 222 SY-----AMQAEKFPSEPNRLLLHGFLENVHFAHTSILSFLVRAG-IPYDLOIYPOER 276
 DB 735 SYHNTAL-----AQANRFLLMHGTGDDNVHFNLSLKFLLDLLNGVENVYDVHVPDS 788
 QY 277 HSIR 280
 DB 789 HSIR 792

RESULT 11

T25173
 hypothetical protein T23f1.7a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T25173
 R.Wilkinson, J.
 submitted to the EMBL Data Library, October 1996

A:Reference number: Z19990
 A:Accession: T25173

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-779 <WIL>
 A:Cross-references: EMBL:Z81129; PIDN:CAB03411.1; GSPDB:GN00023; CESP:T23f1.7a

A:Experimental source: clone T23f1
 C:Genetics:

A:Gene: CESP:T23f1.7a
 A:Map position: 5

A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1

C:Superfamily: dipeptidyl-peptidase IV

Query Match 18.1%; Score 303.5; DB 2; Length 779;

Best Local Similarity 25.7%; Pred. No. 3e-18; Indels 89; Gaps 14;
 Matches 93; Conservative 62; Mismatches 118;

QY 7 SPLEHLLVYVSVNPGVTRLTDRGYS--HSCCISQHCDFISKYSNOKNPHCVSLYKL 63
 DB 431 APKPSHRSLSFSY-----SLADESRNAGYCSISKCTMAQAQMDQMTAIVSKCP 483
 QY 64 SSPED-----DPTCKTK-----EFMATILDSAGPLPDYTPPEIFSESTGFTLY 108
 DB 484 AAPHTAIYVLRMSDDKTEHANLLYDTYQNRVEAG-LPVITIKETI---KISDFDAL 539
 QY 109 GMLYRPHDLPQPKKY---PTVLYFYGPO----- 134
 DB 540 IKLSIPKDIYNRDKQAIPLIVHYVGGNDQNTKATQIGIEVVASASQAAILRIDRG 599
 QY 135 -----GQIEIDDOVEGLQYLASRY-DFIDLRVGIHWSYGYLSLMAI 177
 DB 600 SCGRGMKYSATYGGQITVEVEDQIKAIKVVLRILYRHLDDARRVAVFGMSYGFMTLSMV 659
 QY 178 MORSD-IRRVAIAGAPVTLMEFYDTGTERWGMHPDQEGCYLYGSVAMQAEKFPSEPR 236
 DB 660 NEAPEQFPCAVSAPVPTNFAVYDATERYWG--DAPLESY--SDVTRKLDNFKS--TR 713
 QY 237 LLLHGFLENVHFAHTSILSFLVRAGKPYDLOIYPOERHSIRVSEGEHYELHLHYL 296
 DB 714 LLMHGLDDNVHFNLSALIDELONRGVDDLMYTPQNAHLSRTS--HYVGCMTFL 771
 QY 297 OE 298
 DB 772 RQ 773

RESULT 12

T25174
 hypothetical protein T23f1.7b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T25174
 R.Wilkinson, J.
 submitted to the EMBL Data Library, October 1996

A:Reference number: Z19990
 A:Accession: T25174

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-799 <WIL>
 A:Cross-references: EMBL:Z81129; PIDN:CAB03412.1; GSPDB:GN00023; CESP:T23f1.7b

A:Experimental source: clone T23f1
 C:Genetics:

A:Gene: CESP:T23f1.7b
 A:Map position: 5

A:Introns: 13/3; 52/1; 111/3; 197/3; 366/3; 524/1; 557/1; 648/2; 703/1; 743/1
 C:Superfamily: dipeptidyl-peptidase IV

Query Match

Best Local Similarity 25.7%; Pred. No. 3.1e-18;
 Matches 93; Conservative 62; Mismatches 118; Indels 89; Gaps 14;

QY 7 SPLEHLLVYVSVNPGVTRLTDRGYS--HSCCISQHCDFISKYSNOKNPHCVSLYKL 63
 DB 431 APKPSHRSLSFSY-----SLADESRNAGYCSISKCTMAQAQMDQMTAIVSKCP 503
 QY 64 SSPED-----DPTCKTK-----EFMATILDSAGPLPDYTPPEIFSESTGFTLY 108
 DB 504 AAPHTAIYVLRMSDDKTEHANLLYDTYQNRVEAG-LPVITIKETI---KISDFDAL 559
 QY 109 GMLYRPHDLPQPKKY---PTVLYFYGPO----- 134
 DB 560 IKLSIPKDIYNRDKQAIPLIVHYVGGNDQNTKATQIGIEVVASASQAAILRIDRG 619
 QY 135 -----GQIEIDDOVEGLQYLASRY-DFIDLRVGIHWSYGYLSLMAI 177

Db 620 SGGRGWKYSRAITYGQITVEVEDQIIAIVLVLTLYLHLLDARKVAVAFGMSYGGFMTLSM 6799
 QY 178 MORSD-IRVAIAGAPVTMTMTITDYGTERTYRNGHPQONEGYYLGVAOAEKFPSEPNR 2266
 Db 680 NEAPQIFKCAVSVAVPVTFAYYDATTERTYRNG--DAPLESY--SDVTKLIDMFKS--TR 7333
 QY 237 LLLHGFDENVFAHTSILLSFLVYAGKRPDYQIYPOEHSIRVPESGGEHYELHLLHYL 2366
 Db 734 LLLMHGLDDNVHFPQNSAIIIDELQNRGVDFDLMTVPNQAHSLSRST--HVVGKMTHTFL 7311
 QY 297 QE 298
 Db 792 RQ 793

Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A.Title: Differential expression of two distinct forms of mRNA encoding members of a dipeptidyl-peptidase family
A.Reference number: A41793; MUID:92108018; PMID:1725689
A.Accession: A41793
A.Status: preliminary; not compared with conceptual translation
A.Molecule type: mRNA
A.Residues: 1-803 <MAD>
A.Cross-references: GB:M76429; NID:G408719; PIDN:AAC41623.1; PID:G408720
A.Note: sequence extracted from NCBI backbone (NCBI:75138)
C.Superfamily: dipeptidyl-peptidase IV
C.Keywords: dipeptidylpeptide hydrolase; glycoprotein
E.257,342/Binding site: carboxylate (Asn) (covalent) #status predicted


```
Db      541 DYNLPWQILKPATFTDTTHYPPLLVDGTPGSQVAEKEFEVSWEVTWVSSHGAVVVKCDG 600
Qy      137 -----IEIDPVEGLQYLASRYDFIDDRVGIHGMSYGYLSLMA 176
Db      601 RSGSGOGTKLHVRRLGLLEKXDMEAVRTML-KEQYIDRTRVAVFGDYGGYLSYI 659
Qy      177 LMQRSD----IFRVAIAGAPVTLMIFYDGYTERYMGHPDNEQGYLGSVAMQAEKPPS 232
Db      660 LPAKGENOGOTFTCSALSPITDFKLYASAFSERYLGHLGDNRAVEMTKVAHRVSAL-- 717
Qy      233 EPNRLILLHGFUDENVHFAHTSILSLFVPAKPYDLOIYPOERHSIRVPESGEHYELHL 292
Db      718 EEOQFLIHPTADEKIHFOHTAELITOLIRKANYSIQIYPDESHYFTSSSLKQHLYRSI 777
Qy      293 LHYLOE 298
Db      778 INFVE 783
```

Search completed: October 15, 2003, 17:14:46
Job time : 9.14576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 (Search time 4.9182 Seconds)

(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680

Sequence: 1 FEGTSDSPLEHLYVSYVN.....HLHYLQENIGSRIALAKVI 310

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.5	21.4	793	1	YEAB_SCHPO 014073 schizosacch
2	345.5	20.6	765	1	DPB4_BOVIN P81425 bos taurus
3	344.5	20.5	767	1	DPB4_RAT P14740 rattus norv
4	341.5	20.3	760	1	DPB4_MOUSE P28843 mus musculu
5	335.5	20.0	765	1	DPB4_FELCA Q9217 felis silve
6	333.5	19.9	766	1	DPB4_HUMAN P27487 homo sapien
7	314	18.7	931	1	ST13_YEAST P33694 saccharomyc
8	310	18.5	818	1	DAF2_YEAST P18862 saccharomyc
9	309.5	18.4	760	1	SEPR_MOUSE P97321 mus musculu
10	302	18.0	760	1	SEPR_HUMAN Q12884 homo sapien
11	262	15.6	863	1	DPB6_BOVIN P42659 bos taurus
12	248	14.8	865	1	DPB6_HUMAN P42658 homo sapien
13	244	14.5	853	1	YDZ2_SCHPO Q9218 schizosacch
14	241	14.3	804	1	DPB6_MOUSE P46101 rattus norv
15	240	14.3	859	1	DPB6_RAT P39839 bacillus su
16	152.5	9.1	657	1	YXUL_BACSU P13798 homo sapien
17	140	8.3	732	1	ACPH_HUMAN P13676 rattus norv
18	138	8.2	732	1	ACPH_RAT P19205 mus scrofa
19	126.5	7.5	732	1	ACPH_PIG P34422 caenorhabd
20	112.5	6.7	740	1	YLJ1_CHEEL O29442 archaeoglob
21	104.5	6.2	298	1	Y812_ARCFU O06903 aeromonas h
22	103.5	6.2	689	1	PCB6_AERYH P81171 rickettsia
23	94	5.6	722	1	Y174_RICPR P14688 b gramicidi
24	92.5	5.5	4451	1	GRSB_BACBR P37287 homo sapien
25	91	5.4	484	1	PICB_HUMAN P44945 haemophilus
26	89.5	5.3	1121	1	EXSC_HAEIN P17011 homo sapien
27	88	5.2	1181	1	ITAZ_HUMAN P23133 pseudomonas
28	87.5	5.2	276	1	TODF_PSEPU P40363 saccharomyc
29	87	5.2	299	1	YJG8_YEAST Q30408 b tyrocidin
30	87	5.2	3587	1	TYCB_BREPA P27134 rattus norv
31	86.5	5.1	264	1	COMT_RAT P31320 blastoclad
32	86.5	5.1	403	1	KAPR_BLAEM P38137 saccharomyc
33	86.5	5.1	543	1	PAT2_YEAST

34	86	5.1	352	1	PSBD_SPIOL P06005 spinacia ol
35	86	5.1	353	1	PSBD_ARATH P56761 arabidopsis
36	86	5.1	353	1	PSBD_PINTH P41644 pinus thunb
37	86	5.1	353	1	PSBD_POPDE Q94948 populus del
38	85.5	5.1	286	1	DHA2_MYCTU Q50600 pseudocateri
39	85.5	5.1	371	1	YMI8_PSEAE Q01609 pseudomonas
40	85	5.1	156	1	RNH_ZYMOA Q09014 zymomonas m
41	84	5.0	493	1	PCD_AROXX Q01470 arthroacte
42	83.5	5.0	353	1	PSBD_WHEAT Q36814 triticum ae
43	83.5	5.0	705	1	PCB6_FLAME P27028 flavobacter
44	83.5	5.0	705	1	PCB6_FLAME P27195 flavobacter
45	83	4.9	739	1	BAC1_MOUSE P97302 mus musculu

ALIGNMENTS

RESULT 1
YEAB_SCHPO STANDARD; PRT; 793 AA.

AC 014073;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
GN SPACNKK4.08 OR SPACZB1.08
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1] 9
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras K., Peat N., Hayles J., Baker S., Baaham D., Bowman C.M.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hildbert H.,
RA Borzym K., Langer I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purrelle B.,
RA Goffeau A., Cadieu E., Drenth S., Gloux S., Lelure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucass M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.,
RL The genome sequence of Schizosaccharomyces pombe.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like vacuoles (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC

DR	EMBL; AL031180; CAA20138.1; -	
DR	PIR; T41703; T41703.	
DR	Genedb.Spombe; SPACUNK4_08; -	
DR	InterPro; IPR002469; DPIP_V_N.term.	
DR	InterPro; IPR001375; peptidase_S9.	
DR	InterPro; IPR002471; Prol_endopep_ser.	
DR	InterPro; IPR000379; Ser_ester_site.	
DR	Pfam; PF009310; DPIP_V_N_term; 1.	
DR	Pfam; PF00326; Peptidase_S9; 1.	
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.	
KW	Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;	
KW	Serine protease; Transmembrane; Glycoprotein; Signal anchor.	
KW	CYTOPLASMIC (POTENTIAL).	
FT	DNAIN	1 24
FT	TRANSEM	25 45
FT		
FT	DOMAIN	793
FT	ACT_SITE	647 647
FT	ACT_SITE	722 722
FT	ACT_SITE	755 755
FT	CARBOHYD	101 101
FT	CARBOHYD	136 136
FT	CARBOHYD	246 246
FT	CARBOHYD	299 299
FT	CARBOHYD	303 303
FT	CARBOHYD	324 324
FT	CARBOHYD	336 336
FT	CARBOHYD	377 377
FT	CARBOHYD	384 384
FT	CARBOHYD	407 407
FT	CARBOHYD	535 535
FT	CARBOHYD	761 761
FO	SEQUENCE	793 AA; 91304 MW; 20B70F97FE21463 CRC64;

	Query Match	21.4%	Score 359.5;	DB 1;	Length 793;	
	Best Local Similarity	31.5%	Pred. No. 4.2e-24;			
	Matches 113;	Conservative 43;	Mismatches 128;	Indels 75;	Gaps 17;	
Qy	5	KDSPLEHHLYVSVYNPGEVTRLTDGYS--HSCCISQHCDFISKYSNOKNPCHVS	61			
Db	442	KDS-TERHLVVYS-IDTLEIYGITDNGDEGYSTSPFDDFYLVNHGPDPV---	mQ 495			
Qy	62	KLSSPEDPRCKTEFMATITLDSAGRLDYTPREL---PSSESTGTGLXGM.LKPHDL	117			
Db	496	ELRSTKDKDXYCLSLSETNSRLKQQ--LSSTILPSVEYEGKLTIFNDTT-FNF-MERRPNPF	549			
Qy	118	QRKKKYPTVLRIYGGPOGQ-----TEIDQVEG-----LoY 148				
Db	550	DVNKKYPVLPFRAYGGPGSQVAKLFRVDFOAYLASHPDFFFIIVTLDSRGTFGNMFRY	609			
Qy	149	LASR-----YD-----FTLDKRVGHGMSYGGYLISMALMQRSDFRVAIA	189			
Db	610	SVSRHLGWESYDGOQAQGFADLPFDVENHVHGWSYCGYLLTKTL-ETODVFSYGA	668			
Qy	190	GAPVTLMFUYDTGTERYMGRHPDNOEGUUGSYVAMQEKPESEPENRTLLHGFLDENVH	249			
Db	669	VAPVTDMLRYDSVTTYTERYMDLRQNKEG-YKNSQIHDEKR-KOUKPFVNAHGDDNVH	726			
Qy	250	FAHTSILLISFLVRAG-KPYDIQIYPOEHSIRVPESGHEVLAHLHYLOENIGSRIMAL	307			
Db	727	FQSHMHLMIDGINTLANCYVMADVAFPPDSAHSI----SYNNASTLSYHRLSIEWIGDALGRI	781			
RESULT 2						
DPp4_BOVIN	ID_DPp4_BOVIN	STANDARD;	PRT;	765 AA.		
AC	P81425; O8MMG8;					
DT	15-DEC-1998 (Rel. 37, Created)					
DT	28-FEB-2003 (Rel. 41, Last sequence update)					
DT	28-FEB-2003 (Rel. 41, Last annotation update)					
DE	Dipeptidyl peptidase IV (EC 3.4.14.5) (DBP IV) (T-cell activation antigen CD26) (adenosine deaminase complexing protein) (ADCP-I)					
DE	(Activation molecule 3) (ACT3) (WC10).					
NN	DPp4 OR CD26.					

OS Boe taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN RP SEQUENCE FROM N.A.
RN RC TISSUE=Lymphocytes;
RX MEDLINE=22067734; PubMed=12073152;
RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
RA Bohach G.A.;
RT "Molecular characterization of bovine CD26 upregulated by a
RT streptococcal superantigen.";
RL Immunogenetics 54:216-220(2002).
RN [2]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RN RC TISSUE=Thymus;
RX MEDLINE=22021197; PubMed=11981836;
RA Gliddon D.R., Howard C.J.;
RT "CD26 is expressed on a restricted subpopulation of dendritic cells in
RT vivo";
RL Eur. J. Immunol. 32:1472-1481(2002).
RN [3]
RN RP SEQUENCE OF 1-24.
RN RC TISSUE=T-cell;
RX MEDLINE=21482004; PubMed=11598101;
RA Lee S.-U., Peters W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
RA Naessens J., Bohach G.A.;
RT "Identity of activation molecule 3 on superantigen-stimulated bovine
RT cells is CD26";
RL Infect. Immun. 69:7190-7193(2001).
RN [4]
RN RP SEQUENCE OF 537-546.
RN RC TISSUE=Kidney;
RX MEDLINE=98293306; PubMed=9629661;
RA Ben-Shooshan I., Parola A.H.;
RT "The CP-1 subunit of adenosine deaminase complexing protein from calf
RT kidney is identical to human, mouse, and rat dipeptidyl peptidase
RT IV.";
RL Comp. Biochem. Physiol. 119B:289-292(1998).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline. Binds and regulates the activity
CC of ADA.
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
CC several immune system tissues.
CC -1- PFM: The soluble form (SDP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AF461806; AAL67836.1; -;
CC EMBL, AY056634; AAL23628.1; -;
CC MEROPS, S09_003; -;
CC InterPro, IPR002471; Prol endopep ser.
CC InterPro, IPR000379; Ser esters site.
CC Pfam, PR000930; DDPV_N term; 1.
CC Pfam, PF00326; Peptidase_S9; 1.
CC PROSITE, PS00708; PRO_ENDOPEP_SER; 1.
CC Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
CC

KM Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
 FT CHAIN 38 765 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 629 765 CHANGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 707 707 CHANGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 765 AA; 88369 MW; E32165421F43E116 CRC64;

Query Match 20.6%; Score 345.5; DB 1; Length 765;
 Best Local Similarity 28.9%; Pred. No. 7.1e-23;
 Matches 101; Conservative 46; Mismatches 136; Indels 67; Gaps 10;

QY 5 KDSPLLEHLVYVSVNBEVTRLT---DRGVSHSCCISQHCDFISKYSNOMNPHCVS 59
 422 KGMFGANLVKIQINDYTKVTCSCELNPRCQYVSFSQBAKYVGLRCSGRLP-F 477
 DB 60 LYKLSSEDDPTCKTKFPAFTILDSAGPLPD-YTPPEIFSEFSTGTTL-YGMLYKHDL 117
 478 LYLTHSNNSNKEKRLVE---NNSDLDOVLQDVQWPSKRLFFHLHGKTFYQMILPPH-F 533
 QY 118 QPKKYPVTFIYGGP----- 133
 DB 534 DKSKKPLLELVNAGPSCQADALFRILNMTYLASTNIIIVASFDGSGSYQGDKNHAI 593
 134 ---QGQLEIDQVEGLQYLAARYDFIDLDRVGHGWSYGYLSMALMQRSDIFRVAIAG 190
 DB 594 NRRLGTREVEDQLEATNQF-SKMGFVDDKRIALMGWSYGVVSMVLGASGVFKCIIAV 652
 QY 191 APYTLTIFYDTGTYTTRYMG--HPDQNEQYGLSVAMQAEKFPSEPNRLILHGFLEENV 248
 653 APVSKMEYDVSYTERYMGLPTEPDNLSYRNSTVMSRAENF--KQVEYLLIHGTADNV 710
 DB 249 HFAMHTSLSLFLVRAGKRYDLQYPOBRHSIRVSESEHMLHLHTLOE 298
 711 HFQOSAOISKALVDAGVDPOSMWTTDDHGLASTAHQHTYTHMSHPLKQ 760

RESULT 3
 DPDA RAT STANDARD; PRT; 767 AA.
 AC P14740;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (GPI10 glycoprotein) (Bile canalicular domain-specific membrane glycoprotein).
 GN DPDA OR CD26.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89123496; PubMed=2563382;
 RA Ogata S., Miumi Y., Ikehara Y.;

RT "Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA and identification of the NH2-terminal signal sequence as the membrane-anchoring domain.";
 RT J. Biol. Chem. 264:3596-3601(1989).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88065516; PubMed=3479775;
 RA Hong W., Doyle D.;
 RT "cDNA cloning for a bile canalicular domain-specific membrane glycoprotein of rat hepatocytes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
 RN [3]
 RP SIGNAL-ANCHOR.
 RX MEDLINE=90338089; PubMed=1974258;
 RA Hong W., Doyle D.;
 RT "Molecular dissection of the NH2-terminal signal/anchor sequence of rat dipeptidyl peptidase IV.";
 RL J. Cell Biol. 111:323-326(1990).
 RN [4]
 RP SEQUENCE OF 281-302.
 RC TISSUE=Kidney;
 RX MEDLINE=94128239; PubMed=7905271;
 RA Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;
 RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl peptidase IV.";
 RL Biol. Chem. Hoppe-seyler 374:973-975(1993).
 CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE PENULTIMATE RESIDUE IS PROLINE.
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-Xcc, from a polypeptide, preferentially when Xbb is Pro, provided Xcc is neither Pro nor hydroxyproline.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN A SOLUBLE FORM.
 CC -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP) BY PROTEOLYTIC PROCESSING.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04591; AAA41096.1; -;
 DR EMBL: J02997; AAA41272.1; -;
 DR PIR: A39914; A39914.
 DR MEROPS: S09.003; -;
 DR InterPro: IPR002469; DPPV_N_term.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Prol_endopep_ser.
 DR InterPro: IPR000379; Ser_astri_site.
 DR Pfam: PF00930; DPPV_N_term; 1.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
 KM Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN 1 767 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
 FT CHAIN 37 767 (MDPP).
 FT CHAIN 281 767 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT (SDPP).
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 29 767 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 631 631 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 709 709 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 741 741 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 38 38 A -> R (IN REF. 2).
FT CONFLICT 183 183 I -> T (IN REF. 2).
FT CONFLICT 332 332 T -> N (IN REF. 2).
FT CONFLICT 352 352 C -> V (IN REF. 2).
FT CONFLICT 394 394 V -> D (IN REF. 2).
FT CONFLICT 562 562 L -> F (IN REF. 2).
FT CONFLICT 624 624 R -> Q (IN REF. 2).
SQ SEQUENCE 767 AA; 88003 MM; AA170CECGF084652 CAC64;

Query Match 20.5%; Score 344.5; DB 1; Length 767;
Best Local Similarity 29.0%; Pred. No. 8,7e-23;
Matches 104; Conservative 50; Mismatches 120; Indels 85; Gaps 15;

QY 5 KDSPLEHLYVSVNPGEVRLTDRGYSHSCIS-----OHCPFF--ISKYGNQKNPH 56
D 424 KEMCGGNLYKI-----QLTD--HTNKKCSCDLNPERCQYVSLSKEAKYQLG 472
QY 57 C---VSLYKLSPEDEPCTCKTKEFWATILDSAGPLPDYRPE--IFSFESTGTLYG 109
D 473 CRGGLPLTYLHRSTQKEKLEVED--NSALDKM--LQGVQPSKLDPIVLNERRF-WYQ 528
QY 110 MLYKPHLOQPKKKRYPLFYIGP----- 133
D 529 MLRPPH-FDKSKRYPLLDVYAGPSQKADAFRLNATYLAISTENITVASFDGRSGCYO 587
QY 134 -----OCQIEIDQVEGL-QYLASRYDFIDLRVGHMGSGYSLMALMORS 181
D 588 GDKIMHAINKRLGLVEVDQLEAARQFL--KMGFVDSKRAVIMKMSIGYITSNVLGSGS 645
QY 182 DIFRAVIAQAPVTIWIIFYDGYTERYWG--HPDNGOQYVIGVAMQAEKFPSPNRLLL 239
D 646 GVFCGIAVAVPVSWEYDYSYTERYMWGLPYPEENLDHYRSTVMSRAENF--KQVEYLL 703
QY 240 LHGFLDNVHFANHSILSLFLYRAGKPRDLOIYQOEHRISRVPSGGEYELHLLHYOE 298
D 704 IHGTRADNVHFOOSAISKALVDAGVDFQAMWYDEDEHGASSTAHQHTYHSMHFLOQ 762

RESULT 4
DPP4 MOUSE
ID DPP4 MOUSE STANDARD; PRT: 760 AA.
AC P28843;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
antigen CD26) (thymocyte-activating molecule) (THAM).
GN DPP4 OR CD26.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Thymus;
RX MEDLINE=92129286; Pubmed=1370813;
RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
RA Pierres M.;
RT "cDNA cloning for mouse thymocyte-activating molecule. A
RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT subgroup of serine proteases."
RL J. Biol. Chem. 267:2200-2208 (1992).
RN [2]
RP REVISIONS.
RA Marguet D.A.;

```

```

RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bi0.A; TISSUE=Liver;
RX MEDLINE=95092780; Pubmed=7999781.
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
RL Biochemistry 33:15204-15214 (1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,
RA Stachenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshylyuk S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faneay J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=91302787; Pubmed=1712807;
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA Bernard A.-M., Gorvel J.-P., Pierres M.;
RT "Evidence that thymocyte-activating molecule is mouse CD26
RT (dipeptidyl peptidase IV).";
RL J. Immunol. 147:447-454 (1991).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline.
CC CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form.
CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X58384; CA041274.1; -
DR EMBL; U12620; AAA82213.1; -
DR EMBL; U12599; AAA82213.1; JOINED.
DR EMBL; U12600; AAA82213.1; JOINED.
DR EMBL; U12601; AAA82213.1; JOINED.
DR EMBL; U12602; AAA82213.1; JOINED.
DR EMBL; U12603; AAA82213.1; JOINED.
DR EMBL; U12604; AAA82213.1; JOINED.
DR EMBL; U12605; AAA82213.1; JOINED.
DR EMBL; U12606; AAA82213.1; JOINED.
DR EMBL; U12607; AAA82213.1; JOINED.
DR EMBL; U12608; AAA82213.1; JOINED.
DR EMBL; U12609; AAA82213.1; JOINED.

```

```

DR EMBL; U12610; AAA82213.1; JOINED.
DR EMBL; U12611; AAA82213.1; JOINED.
DR EMBL; U12612; AAA82213.1; JOINED.
DR EMBL; U12613; AAA82213.1; JOINED.
DR EMBL; U12614; AAA82213.1; JOINED.
DR EMBL; U12615; AAA82213.1; JOINED.
DR EMBL; U12616; AAA82213.1; JOINED.
DR EMBL; U12617; AAA82213.1; JOINED.
DR EMBL; U12618; AAA82213.1; JOINED.
DR EMBL; U12619; AAA82213.1; JOINED.
DR EMBL; BC022183; AAA82213.1; JOINED.
DR MEROPS; S09.003; -.
DR MGD; MGI:94919; DPP4.
DR InterPro; IPR002469; DPP4V N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estcs_site.
DR Pfam; PF00930; DPP4V_N_term; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
DR HydroLase; Aminoacylase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 760
FT CHAIN DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT CHAIN (MDP).
FT CHAIN DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT CHAIN (SDP).
FT CHAIN CYTOPLASMIC (POTENTIAL).
FT CHAIN SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT CHAIN (POTENTIAL).
FT CHAIN EXTRACELLULAR (POTENTIAL).
FT CHAIN CYS-RICH.
FT CHAIN CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHAIN CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHAIN ACT_SITE 702 702
FT ACT_SITE CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 83 83
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 213 213
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 514 514
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 760 AA; 87436 MW; A5F644B46E4A3DF8 CRC64;

```

```

Query Match 20.3%; Score 341.5; DB 1; Length 760;
Best Local Similarity 28.0%; Pred. No. 1.6e-22;
Matches 101; Conservative 51; Mismatches 120; Indels 89; Gaps 14;

```

```

QY 5 KDSPLHLLVSVYVNGEVRRLTRDGYSHSCIS-----OHCDPFIKYNOKNPH--- 56
DB 417 KEMPGGRNLYKI-----QLTID--HNWVKLSCDLNPCKQYAVSFSEAKYQUG 465
QY 57 C---VSLYKSSPEDDPTCKTEKEMATILDSAG---PLDYTPPE---IFSEESTTGT 106
DB 466 CMGPGLEPLTTLHRSSTDHKEKRLVLE-----DNSALDRMDQVMPKSKLDIFLNETRF- 518
QY 107 LYGMLYKPRDLPGRKKYPTVLFYIGP----- 133
DB 519 WQOMLTPPH-FPKSKYPLLLVDYAPCSQKADASRLMWATYLASTENIIVASFDGRGS 577
QY 114 -----OQOIEIDDOVEGLQYLASRYDFIDLDRAVGHGMSYGYLSTLALMQ 179
DB 578 GYGQDKIMHAINRRLGTELEVEDOIEAARQFV-KMGVVDKRAVIMKMSYGYVTSNVLGS 636
QY 180 RSDITRVAIAGAPVTLMTFYDTGTERYNG--HPDNEGGYILGSAVMAQEKPPSEPNRL 237
DB 637 GSGVFCGGAIVAPVRMEYTDVSYTERYVGLPLPEPNLDHYNSTYMSRAEHF--KQVEY 694
QY 238 LLLHGFLENDVFAHTSILSLFVLRAGKPYDQLQYPOEHSIRVPSGSHYELHLHYAQ 297
DB 695 LLIHGTADNVHFFQOASAQISKALVDAGVDFOAMWYTDDEHGASSTAHQHYTHSHSHFLQ 754

```

```

QY 298 E 298
DB 755 Q 755

RESULT 5
DPP4_FELCA STANDARD; PRT; 765 AA.
ID DPP4_FELCA
AC 09N217;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxId=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripherel blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue."
RL Immunogenetics 50:366-368(1999).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family 59B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB023954; BAA92344.1; -.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPP4V N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estcs_site.
DR Pfam; PF00930; DPP4V_N_term; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW HydroLase; Aminoacylase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765
FT CHAIN DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT CHAIN (MDP).
FT CHAIN DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT CHAIN (SDP) (BY SIMILARITY).
FT CHAIN CYTOPLASMIC (POTENTIAL).
FT CHAIN SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT CHAIN (POTENTIAL).
FT CHAIN EXTRACELLULAR (POTENTIAL).
FT CHAIN CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 629 765
FT ACT_SITE CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 707 707
FT ACT_SITE CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 84 84
FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91

```



```

FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 88213 MW; 3EFC98A22817509 CRC64;

Query Match 20.0%; Score 335.5; DB 1; Length 765;
Best Local Similarity 27.5%; Pred. No. 5.5e-22;
Matches 98; Conservative 53; Mismatches 128; Indels 77; Gaps 14;

OY 12 HLTVVSVVNGP-----EVTRLDGRGSHSCIS-----QHCDPEFISYNSQKPH---C- 57
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 413 YLYTISNEYKMGNGNLYKIQANDYTKVACLSCELPKRCQYYSFSREAKTYQLRCS 472
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 58 ---VSLYKLSPPDDPTCKTKFEWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGML 111
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 473 GPCLPLTYLHRSNDEBLRYLED--NSALDKM--LQEVOMPSKKLDLILNETKF--WYQMT 528
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 112 YKPHDQPKKYTYVLEFYGSP----- 133
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 529 LRPFH-FDTSKRYPLLDVYAGPSCQKADALFRLNMTATYSTENIIVASFDRGSGYQGD 587
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 134 -----QGQIEIDDOVEGLQYLARSYDFIDLDRVGHIGMSYGYLSTLMALMQRSDIF 184
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 588 KIMHAYNRRLGTEVEVDQIEANQF--SKMGFVVDKRIALIMGWSYGYVSMVLAGSGVF 646
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 185 RVAIAGAPVTLWIFDTGYTERYMGHP--DNEQGYVIGSVAM--QAEKPEPSEPRLLLLG 242
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 647 KCGIAVAPRSMRYSVYTERYMGPLPTQPDNLDYKNSITVMSRAENF--KQVEYLLING 704
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 243 FLIDENHFAHTSTLLSFLVAGKRYDQIYPOERHSIRVSESEHNEHLNHTLOE 298
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 705 TADDNVHFOQSAQISKALVDAGVDFQAMWYTTDEDHGIASGPAHQHLYTHMSHPKIQ 760
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 6
DPP4 HUMAN STANDARD; PRT; 766 AA.
ID _DPP4_HUMAN
AC P27487;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
antigen CD26) (TSP103) (Adenosine deaminase complexing protein-2)
DE (ADAP).
GN DPP4 OR ADCP2 OR CD26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=92329551; PubMed=1352704;
RT Mismata Y., Hayashi Y., Arakawa F., Ikehara Y.;
RT "Molecular cloning and sequence analysis of human dipeptidyl
peptidase IV, a serine proteinase on the cell surface.";
RL Biochim. Biophys. Acta 1131:333-336(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA MEDLINE=95012454; PubMed=7927537;
RT Abbot C.A., Baker E., Sutherland G.R., McCaughan G.W.;
RT "Genomic organization, exact localization, and tissue expression of
the human CD26 (dipeptidyl peptidase IV) gene.";
RL Immunogenetics 40:331-336(1994).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Peripheral blood;
RX MEDLINE=92325476; PubMed=1352530;
RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
RA Dahlberg H.N., Schlossman S.F., Morimoto C.;
RT "Cloning and functional expression of the T cell activation antigen
CD26.";
RL J. Immunol. 149:481-486(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=93171637; PubMed=8094732;
RA Tanaka T.;
RL J. Immunol. 150:2090-2090(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Smetten C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.C.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinaki M.I., Skalska U., Smalhus D.E.,
RA Schenker A., Schein J.E., Jones S.J.M., Marx M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 1-551 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=92165847; PubMed=1347043;
RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA Trocet P., Barbat A.;
RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
human coding sequence and changes of dipeptidyl peptidase IV mRNA
levels during cell differentiation.";
RL J. Biol. Chem. 267:4824-4833(1992).
RN [7]
RP SEQUENCE OF 545-766 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=91024044; PubMed=1977364;
RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
RT "Isolation of a cDNA probe for the human intestinal
dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
chromosome 2.";
RL Ann. Hum. Genet. 54:191-197(1990).
RN [8]
RP SEQUENCE OF 1-31 FROM N.A.
RX MEDLINE=96067599; PubMed=7487939;
RA Boehm S.K., Gum J.R., Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
regulation from a TATA-less GC-rich sequence characteristic of a
housekeeping gene promoter.";
RL Biochem J. 314:835-843(1995).
RN [9]
RP PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=93210468; PubMed=8096237;
RA Morrison M.E., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA Houghton A.N.;
RT "A marker for neoplastic progression of human melanocytes is a cell
surface ectopeptidase.";
RL J. Exp. Med. 177:1135-1143(1993).

```

CC - FUNCTION: Removes N-terminal dipeptides sequentially, from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline. Plays a role in T cell activation.
 CC - CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-
 CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC Xcc is neither Pro nor hydroxyproline.
 CC - SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
 CC - SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form.
 CC - PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing.
 CC - SIMILARITY: Belongs to peptidase family S9B.
 CC - DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U13735; AAB60646.1; -;
 CC DR EMBL, U13710; AAB60646.1; JOINED.
 CC DR EMBL, U13711; AAB60646.1; JOINED.
 CC DR EMBL, U13712; AAB60646.1; JOINED.
 CC DR EMBL, U13713; AAB60646.1; JOINED.
 CC DR EMBL, U13714; AAB60646.1; JOINED.
 CC DR EMBL, U13715; AAB60646.1; JOINED.
 CC DR EMBL, U13716; AAB60646.1; JOINED.
 CC DR EMBL, U13717; AAB60646.1; JOINED.
 CC DR EMBL, U13718; AAB60646.1; JOINED.
 CC DR EMBL, U13719; AAB60646.1; JOINED.
 CC DR EMBL, U13720; AAB60646.1; JOINED.
 CC DR EMBL, U13721; AAB60646.1; JOINED.
 CC DR EMBL, U13722; AAB60646.1; JOINED.
 CC DR EMBL, U13723; AAB60646.1; JOINED.
 CC DR EMBL, U13724; AAB60646.1; JOINED.
 CC DR EMBL, U13725; AAB60646.1; JOINED.
 CC DR EMBL, U13726; AAB60646.1; JOINED.
 CC DR EMBL, U13727; AAB60646.1; JOINED.
 CC DR EMBL, U13728; AAB60646.1; JOINED.
 CC DR EMBL, U13729; AAB60646.1; JOINED.
 CC DR EMBL, U13730; AAB60646.1; JOINED.
 CC DR EMBL, U13731; AAB60646.1; JOINED.
 CC DR EMBL, U13732; AAB60646.1; JOINED.
 CC DR EMBL, U13733; AAB60646.1; JOINED.
 CC DR EMBL, U13734; AAB60646.1; JOINED.
 CC DR EMBL, M74777; AAA51943.1; -;
 CC DR EMBL, BC013329; AAH13329.1; -;
 CC DR EMBL, M80536; AAA52308.1; -;
 CC DR EMBL, X60708; CAA43118.1; -;
 CC DR EMBL, S79876; AAB55614.1; -;
 CC DR PIR, S24313; CDH026.
 CC DR PDB, IN1M; 07-JAN-03.
 CC DR MEROPS; S09.003; -;
 CC DR GeneW; HGNC:3009; DPP4.
 CC MIM; 102720; -;
 CC DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; TAS.
 CC DR GO; GO:0005209; P:plasma protein; TAS.
 CC DR GO; GO:0006955; P:immune response; TAS.
 CC DR InterPro; IPR002469; DPP4V_N term.
 CC DR InterPro; IPR001375; Peptidase_S9.
 CC DR InterPro; IPR002471; ProL_endopep_ser.
 CC DR InterPro; IPR000379; Ser_eactr_site.
 CC DR Pfam; PF00930; DPP4V_N term; 1.
 CC DR Pfam; PF00326; Peptidase_S9; 1.
 CC DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
 CC DR Hydrolase; Aminopeptidase; dipeptidase; Serine protease;
 CC Transmembrane; Glycoprotein; Signal-anchor; 3d-structure;
 CC CHAIN 1 766
 CC (MDPP).

FT CHAIN 0 39 766 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
 FT DOMAIN 1 6 (SDPP).
 FT TRANSMEM 7 28 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 29 766 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT ACT SITE 630 630 (POTENTIAL).
 FT ACT SITE 708 708 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 740 740 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 6 6 K -> R (IN REF. 6).
 FT CONFLICT 7 7 V -> I (IN REF. 1).
 FT CONFLICT 437 437 S -> I (IN REF. 1).
 FT CONFLICT 557 557 T -> I (IN REF. 7).
 FT CONFLICT 663 663 D -> E (IN REF. 7).

Query Match 19.9%; Score 333.5; DB 1; Length 766;
 Best Local Similarity 27.7%; Pred. No. 8,3e-22;
 Matches 99; Conservative 52; Mismatches 124; Indels 83; Gaps 14;

QY 5 KQSPLEHLLVYVSYNPGVETRLTDGYSHSCIS-----QHCDFFISKYSNOKNPA--- 56
 DB 423 KGMFGGRNLKYI-----QLSD--YTKVTLCSLNPBERQYYSVSFSKAKYQLR 471
 QY 57 C-----VSLKLSPEDDPTCKTEFATILDSAGPLPDTTPPE---IFSESTTGFTLYG 109
 DB 472 CGSPGLPYTLTSHSVNDKGLRVLED--NSALDKN--LQNVQMPKSLDFTILNETKF-WYQ 527
 QY 110 MLKYPHDLQPKKYPVTLFIYGGP----- 133
 DB 528 MLPPH-PDKSKKYPILLDQVAVPCQKADTYFRILMATTYSTETIIYASFGRRSGYQ 586
 QY 134 -----OQOIEIDQVGLQYLASRYFDIDLDVGVHSGSYGYGLSLMALMQRSD 182
 DB 587 GPKIMHAINRRLGTFEVEDQIEAARQF-SKMGVVDNKRILAWMSYGVYTVSWLSSGSG 645
 QY 183 IRRVAAGAPVTLMTFYDGYTERYWG--HPDNEGGYLGSYAMAEKFPSPENRLL 240
 DB 646 VPKGIAVAPVSRWEYDVTERTYWGTPPEPNDLDHYNSTVMSRAENF--KQVEYLLI 703
 QY 241 HGFLENVHFATSTILSFLVAGKPYDQIYPOEHSIRVPESGHEYLHLHYLOE 298
 DB 704 HGTADNVHFFQSAQISKALVDVGVDFQAMWYTTDEDHGIASSTAHQHITYTHMSHTKQ 761

RESULT 7
 ST13 YEAST STANDARD; PRT; 931 AA.
 AC P33854; 01-FEB-1994 (rel. 28, Created)
 DT 01-FEB-1994 (rel. 28, Last sequence update)
 DT 28-FEB-2003 (rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DBAP A) (YSCIV).
 GN STE13 OR YC11 OR YOR219C OR YOR50-9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCB1_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9506382; PubMed=7975897;
 RA Anna-Artiola S.S., Herskowitz I.;
 RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
 RT aminopeptidase.";
 RT Yeast 10:801-810(1994).

```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB320;
RA Flanagan C.A., Thorne J.;
RT "STE13."
RL (in) Getting M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.) ;
RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RN Press, Oxford (1994).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96437977; PubMed=8840505;
RA Galissoon F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae."
RL Yeast 12:877-885(1996)
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/annouce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, L21944; AAA35119.1; -
DR EMBL, U08230; AAA17897.1; -
DR EMBL, X92441; CAA63182.1; -
DR EMBL, Z75127; CAA99437.1; -
DR PIR, A49737; A49737.
DR MEROPS, S09.005; -.
DR SGD, S0005745; STE13.
DR GO, GO:0005802; C:Golgi trans-face; IDA.
DR GO, GO:0004177; F:aminopeptidase activity; IDA.
DR GO, GO:0007323; P:peptide pheromone maturation; IDA.
DR InterPro: IPR002469; DEPIV_N_term.
DR InterPro: IPR001375; Peptidase S9.
DR InterPro: IPR002471; Prol endopep ser.
DR InterPro: IPR000379; Ser_catr_site.
DR Pfam, PF00930; DEPIV_N_term; 1.
DR Pfam, PF00326; Peptidase_S9; 1.
DR PROSITE, PS00708; PRO_ENDOPEP_SER; FALSE_NEG.
KW Hydrolyase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT DOMAIN 1 119
FT TRANSMEM 120 140
FT DOMAIN 141 931
FT ACT_SITE 785 785
FT ACT_SITE 863 863
FT ACT_SITE 896 896
FT CARBOHYD 377 377
FT CARBOHYD 814 814
SQ SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;
Query Match 18.7%; Score 314; DB 1; Length 911;
Best Local Similarity 25.6%; Pred. No. 5.9e-20;
Matches 100; Conservative 54; Mismatches 133; Indels 104; Gaps 12;
1 FEGTKDSELEHLLVYVYV-----NPGEVRLTDRGVSHCSCISQHOQDFISK 48
Db FTANEIGMSQHLVSLTLDSTQNTQPSLONP-----SKYDFIDELSSARVATSK 615
49 YSNQKNP-----HCVSLYKLSPEPDPTCKTKERFWATILDSAGPLPDYT 92
Db KLGDDPTFKVAGPLTRVNLVAEIHDDSLITLKDE-----KFKYD-----KIKYD 660
93 PPELFSEFSTT---GFTLYGMALYKRDHQPCKKYPTVLFYIGGPGQ----- 136

```

```

Db      661 LP-ITSYKTMWLDGVEINVIETKPNALNPKKKXPLVINIGGSGSQFTTKSLAFEGA 719
Qy      137 -----IEIDQVGG-----LQYLAISYDFIDLDLV 161
Db      720 VSGLDVIVIQIIEPRGGTGKMGKMSFRSMAREKLGWEPBDITVEYTKKPIQRNSQIHDESKI 779
Qy      162 GIGHMSYGGYVSLMAL-MORSDFRVAIAGAPVLMIFDYGTERTWGHDPQEGGYL 220
Cc      780 AINWMSYGGFTSLKTVLELDNGDFTFKMAAVALPNTMLLYDSVYTERMTNQGSEKHGEGFE 839
Db      221 GSVAMQAEKPPSEPNRLLLHGFIDENVHFAHTSILISFLVRAG-KPYDQIYQERHSI 279
Cc      840 VSTIQNKSFSES-LKRFIFVHGFTEDDVNHIQNTFRVLVDQMLGLTNYDMHI FPDSDHSI 898
Qy      260 RVPESEGHVELHLHYLOENLGSRIALKYI 310
Db      899 RYHNAQRIVQKLYTMLDAFAERFDNTEVL 929

RESULT 8
DAP2_YEAST
ID_DAP2_YEAST STANDARD; PRT; 818 AA.
AC PI96F2;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPA2 OR YSCV).
GN DAP2 OR YHR028C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=89174971; PubMed=2647766;
RA Roberto C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RT "Structure, biosynthesis, and localization of dipeptidyl
RT aminopeptidase B, an integral membrane glycoprotein of the yeast
RT vacuole.";
RL J. Cell Biol. 108:1363-1373(1989).
[2]
RN RN
RP SEQUENCE FROM N.A.
RX STRAIN=5288C / AB972;
RA MEDLINE=94378003; PubMed=8091229;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Lareelle P., Louis E.J., Macri C., Maris E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevaaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterson R., Wilson R.,
RA Valdin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
Cc -1 SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
Cc VACUOLUS.
Cc -1 SIMILARITY: Belongs to peptidase family S8B.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb.ch/announce/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc DR EMBL; X15484; AAA3512.1; -.
Cc DR EMBL; U10399; AAB6879.1; -.
Cc DR PIR; S46780; A30107.
Cc MEROPS; S09_006; -.
Cc DR SCD; S0001070; DAP2.
Cc GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.
Cc InterPro; IPR002469; DPPIV_N_term.

```

```

DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; Prol_doped_ser.
DR InterPro: IPR000379; Ser_gstrs_site.
DR Pfam: PF00330; DpIv_N-term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
DR Hydrolase: Aminopeptidase; Dipeptidase; Serine protease;
KM Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 30 45 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT 46 818 LOMENAL (POTENTIAL).
FT ACT SITE 679 756 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 756 766 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 769 789 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 43 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 738 738 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 83 83 Q -> H (IN REF. 1).
FT CONFLICT 125 125 S -> N (IN REF. 1).
FT CONFLICT 152 152 FEIGN -> LRLET (IN REF. 1).
FT CONFLICT 182 200 D -> N (IN REF. 1).
FT CONFLICT 200 200 TSINVENESS -> DKRKEKEF (IN REF. 1).
FT CONFLICT 366 375 ARRAFGQFVK -> QSVLSMGNLTNLTLYSSHRDIHKT
FT CONFLICT 808 818 PSYLTMYI (IN REF. 1).
SQ SEQUENCE 818 AA; 93404 MW; 318F45045375BD3 CRC64;

Query Match 18.5%; Score 310; DB 1; Length 818;
Best Local Similarity 26.9%; Pred. No. 1,1e-19;
Matches 98; Conservative 38; Mismatches 122; Indels 106; Gaps 13;

QY 1 FEGRDSPLEHLLVYVSVNPGVETRLDRGVS--HSCCISQHCDFPISKYSNOKNFH-- 56
DB 451 FISTKKSSTERYHYIYDLSRPNBIIIEVDTSEGDYVDFSSGRRRGILLTYGPKVPYVK 510
QY 57 -----C-----VSLYKSSPPDDPTCKTKERWATILDSAGLPPTYPEIFS 98
DB 511 IVDPHSRKAECDCGNVLGKSLYHLEKNE-----VLTKI--LEDYAVPRKSF 555
QY 99 FESTTGFTLYG--MLYKPHDLP-----GKKYPTVLFIYGPP----- 134
DB 556 RELNLGKDFEFGKQILVNSYELLPNDFELSHYFVFFAYAGSPNSQVVKTFSSGFENEV 615
QY 135 -----GQLEIDDOVEGLQYLASRYDFIDLRVG 162
DB 616 VASQLNATVYVVDGRGTGFKGQDFRSLVDRGLDGYARQDIAASLYGS-LTFVDPQKIS 674
QY 163 IHGMSYGYGLSLMALMQRSD-IFRVAIAGAPVTLWTFYDGYTERYMGHPDNEQGYLYG 221
DB 675 LFGMSYGYGLTLTKLEKDGSRHFKYGMVAPTDMFYSYTERYMHFPQENFGYVBS 734
QY 222 SY-----AMQAEKFPSEPNRLLLHGFLENNVFAHHSILSLFLVAG-KPYLQIYPOQR 276
DB 735 SYHNTAL-----AQNRFLLMHGTGDNVHFPQNSLKLFLDLLDNGENVYDVHVPDSD 788
QY 277 HSIR 280
DB 789 HSIR 792

```

```

DE membrane serine protease).
GN PAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_Taxid:10090;
RX SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC STRAIN-BALB/c; TISSUE=Embryo;
RX MEDLINE=97284459; PubMed=9139873;
RA Niedermeyer J., Scanlan M.J., Garin-Chesa P., Daiber C., Fiebig H.H.,
RA Old L.J., Rettig W.J., Schnapp A.;
RA "Mouse fibroblast activation protein: molecular cloning, alternative
RT splicing and expression in the reactive stroma of epithelial
RL cancer.";
RN Int. J. Cancer 71:383-389 (1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feringold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M., Usdin I.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak G.A., Mewan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=P97321-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97321-2; Sequence=VSP_005366;
CC Name=3;
CC IsoId=P97321-3; Sequence=VSP_005369;
CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC embryos from day 7-19 and in new-born mice (P1).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: Y10007; CAA71116.1; -.
CC DR EMBL: BC019190; AAH19190.1; -.
CC DR MEROPS: S09.007; -.
CC MGD: MGI:109608; Fap.

```

```

DR InterPro: IPR002469; DPPIV_N term.
DR InterPro: IPR001375; Peptidase S9.
DR InterPro: IPR002471; Prol endopep ser.
DR InterPro: IPR000379; Ser_ester_site.
DR Pfam: PF00930; DPPIV_N term; 1.
DR Pfam: PF00326; Peptidase S9; 1.
DR PROSITE: PS00708; PRO-ENDOPEP-SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 26 761 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLC 31 35 Missing (in isoform 2).
FT VARSPLC 31 63 /FTID=VSP_005368.
FT VARSPLC 31 63 Missing (in isoform 3).
FT CONFLICT 737 737 S -> L (IN REF. 2).
FT SEQUENCE 761 AA; 87944 MW; 9174CA3EDA213B25 CRC64;
Query Match 18.4%; Score 309.5; DB 1; Length 761;
Best Local Similarity 25.6%; Pred. No. 1.1e-19;
Matches 91; Conservative 59; Mismatches 133; Indels 73; Gaps 10;
OY 1 FEGTDSPLFHLHYVSYNVPGEVTLTDRGYSHCSCISQH-----CDFEIKYSNOKNP 55
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 415 FBG---YPERNNIYKISIGNSPPSK-----CVCCHLRKRCQYTTSPSKAKY 461
56 HCVSLY---KLSPEDPTCKTKFEMATILDSAGLPDYPPEI-FSFESTGTFTLYGM 110
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 462 VALVCYCGPLPISTLHDGRDDEIQVLEENKELENSLRNIQPKVYIKLXKQGLTFWYK 521
OY 111 LYKPHLDQKQKPYVLFYIGGPGQ-IEIDQVDELQTLAR----- 152
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 522 MLPPQFDRSKKPYPLLIQVYGGPCSQSVAFVNMVITLAKREGIVTLVDGRGTAFOG 581
153 -----YD-----FIDLDRVGIGHSGSYGLSLMLMGRSDIF 184
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 582 DKFLNAVYRKLGVEVEDQLTAVRKTIEMGFIDEERIALWMSYGGYSLASGTGLP 641
OY 185 RVAIAGAPVTLWIFYDTGYTERYMGHP--DQNEQGYLGSVAMQAEKPPSEPNRLLLHG 242
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 642 KCGIYAVPVSWEYVYASISERFMGLPTKODNLEHKNSTVVARAEYFRNV--YLLING 699
OY 243 FIDENVHFAHTSILSLFVIRACKPYDQIYPOERHSIRVPEGHEHLHLLTYLQE 298
| | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 700 TADNVHFFONSQIYAKALVNAQVDFQAMWYSQDNHGISGSRQNHLYTHMTHFLKQ 755

```

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fibroblast;
RX MEDLINE=94261645; PubMed=7911242;
RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncasi M.P.,
RT Healey J.H., Old L.U., Rettig W.U.;
RT "Molecular cloning of fibroblast activation protein alpha, a member of
RT the serine protease family selectively expressed in stromal
RT fibroblasts of epithelial cancers.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Melanoma;
RX MEDLINE=97388251; PubMed=9247085;
RA Goldstein L.A., Gherzi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,
RA Plessate D., Chen W.-T.;
RT "Molecular cloning of seprase: a serine integral membrane protease
RT from human melanoma.";
RL Biochim. Biophys. Acta 1361:11-19(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP 511-518.
RC TISSUE=Melanoma;
RX MEDLINE=97218181; PubMed=9065413;
RA Pineiro-Sanchez M.L., Goldstein L.A., Dodd J., Howard L., Yeh Y.,
RA Chen W.-T.;
RT "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT (seprase) as a serine integral membrane protease.";
RL J. Biol. Chem. 272:7595-7601(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Melanoma;
RX MEDLINE=20112818; PubMed=10644713;
RA Goldstein L.A., Chen W.-T.;
RT "Identification of an alternatively spliced seprase mRNA that encodes
RT a novel intracellular isoform.";
RL J. Biol. Chem. 275:2554-2559(2000).
RN [5]
RP SEQUENCE OF 192-208; 220-240 AND 510-521.
RX MEDLINE=94327249; PubMed=7519584;
RA Rettig W.U., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA Garin-Chesa P., Healey J.H., Old L.U.;
RT "Fibroblast activation protein: purification, epitope mapping and
RT induction by growth factors.";
RL Int. J. Cancer 58:385-392(1994).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and may contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC surface lamellipodia, invadopodia and on shed vesicles.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1; Synonyms=L;
CC IsoId=Q12884-1; Sequence=Displayed;
CC Note=Major isoform;
CC Name=2; Synonyms=S, Truncated;
CC IsoId=Q12884-2; Sequence=VSP_005367;
CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC during development, tissue repair, and carcinogenesis.
CC -1- PTM: N-glycosylated.
CC -1- PTM: The N-terminus may be blocked.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U09278; AAB49652.1; -
 DR EMBL; U76833; AAC51668.1; -
 DR EMBL; AF007822; AAR21600.1; -
 DR MEROPS; S09_007; -
 DR Genew; HGNC:3590; FAP.
 DR MIM; 600403; -
 DR InterPro; IPR002469; DPPV_N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR002471; Prol endopep ser.
 DR InterPro; IPR000379; Ser esters site.
 DR Pfam; PF00930; DPPV_N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 DR Prosite; PS00708; PRO_ENDOPEP_SER; 1.
 DR Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Alternative splicing.
 FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 26 760 EXTRACELLULAR (POTENTIAL).
 FT ACT SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 1 521 Missing (in isoform 2).
 FT FTID=VSP_005367.
 FT P -> A (IN REF. 2).
 FT K -> T (IN REF. 2).
 FT R -> T (IN REF. 2).
 FT CONFLICT 207 207
 FT CONFLICT 229 229
 FT CONFLICT 354 354
 SQ SEQUENCE 760 AA; 87820 MW; AAD34B4801BE07EA CRC64;

Query Match 18.0%; Score 302; DB 1; Length 760;

Best Local Similarity 25.5%; Pred.No.5.3e-19;
 Matches 93; Conservative 55; Mismatches 139; Indels 78; Gaps 11;

QY 1 FEGTQDSPLEHLYVSVVPGEVRLTDRGYSHS-CCISQH-----CDFEIKSYSNOKN 54
 DB 401 FRTOQSLFSSNSPEFEYRPRRNIVNISIGSYPPSKKCVCHLXKRCQYVYASFSFDYAK 460
 QY 55 PHCVSLY---KLSPEDDPTCKEFAWATILDSAGLPD-----YTPPEIFSFESITGF 105
 DB 461 YVALVCYGGPISTLHGRTRDQ---IKILEENKELEALKNIOLEPKKEIKLEVDI 516
 QY 106 TLVGMLYKHHDLOPGKKYPTVLFLIYGGP----- 133
 DB 517 TLVYKMLLPQGFDRSKKYPILLVYGGPQSGSVSVFAVNMISYLASKEGMVALVDGRG 576
 QY 134 -----OGQIEIDQVEGLQYLASRYFDLDRYGIHGSYGYLSMALM 178
 DB 577 TAFQSKLXAYVRKLGVEVEEQITAVRKFI-EMGFIDSKRLAIMGWSGGVYSSIALA 635
 QY 179 QRSDFRVAIAGAPVTLMIFYDTGYTERYVGNP--DQNEGGYLVGVAQAQKFPSEPNR 236
 DB 636 SGTGLFKCGIAVAPSVSWEYVASVTERFWGLPTKODNLHYNKSTVMARAEFRVND-- 693
 QY 237 LLLHGFDENVHFAHTSILSLFLVRAKRYDQIYPOEHNSIRVBSG---EHYELHL 293
 DB 694 YLLIHETADNVHFOHNSAQIAKALVNAQVDFQAMWYSDQNHGL-----SGLSHLYTHMT 749
 QY 294 HYLQE 298
 DB 750 HFLKQ 754

RESULT 11
 DPP6_BOVIN

ID DPP6_BOVIN STANDARD; PRT; 863 AA.

DT PA2659;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).

GN DPP6.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;
 RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RA MEDLINE=92108018; PubMed=1729689;
 RA Mada K., Yokotani N., Hunter C., Doi K., Wenholt R.J., Shimasaki S.;

RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).

CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein (probable).
 CC Event=Alternative products;
 CC Name=Dppx-L;
 CC IsoId=P42659-1; Sequence=Displayed;

CC Name=Dppx-S;
 CC IsoId=P42659-2; Sequence=VSP_005364;
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN

CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M76428; AAC41622.1; -
 CC EMBL; M76429; AAC41623.1; -
 CC PIR; A41793; A41793.

DR MEROPS; S09_973; -
 DR InterPro; IPR002469; DPPV_N term.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser esters site.

DR Pfam; PF00930; DPPV_N term; 1.
 DR Pfam; PF00326; Peptidase_S9; 1.
 KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.

FT DOMAIN 1 93 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 94 114 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).

FT DOMAIN 115 863 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 1 79 MASIVRFTGKINTSRSPAPPSARLLGQGGPDEGAPK
 FT PIGAQAAPAPRERGGAAGGRFRFOYQARSQDDED ->
 FT MTTAKPNSGKSVQOQEQ (in isoform DPPX-S).
 FT FTID=VSP_005364.

SQ SEQUENCE 863 AA; 96556 MW; 23DBA792B841A3D CRC64;
 Query Match 15.6%; Score 262; DB 1; Length 863;

Best Local Similarity 25.0%; Pred. No. 2.3e-15;
Matches 89; Conservative 45; Mismatches 152; Indels 70; Gaps 9;

CC
DB 1 FEETKDSPLEHNLVSVNPGVETRL-----LTDKRGVSHSCCISQHCDFPISYKSNQK 54
DB 500 FLSTEDLPKRRRLYSASTV--GSNRCQSCDLVDNCTYSASGSPADPFLKCGEPGV 557
DB 55 PHCVSLYKSSPEDPTCKTKE--FWATILDSAGPLPDPYTPPEIFSFESTGTFLYGLMYK 113
DB 558 P-TVSANTTDKKMFDLETNEHQKALSDRQMKVEYRIE-----TDDYNLPIDLK 610
DB 114 PHDLQPKKYPYVLFYGGP----- 133
DB 611 PATFTDPAHPPLLVVGTGSGSVAEKFAVTWETWVSSHGAVVVKDGRSGFGQTRL 670
DB 134 -----QGQIEIDVOEGLOYLASRYDFIDLVDVGHGMSYGGYLSLMLMRSD----- 182
DB 671 LHEVRRRLGSLSEEDQMEAVRWML--KEPYIDKTRVAVFGDYGGLTYLLPAKGDQAP 729
DB 183 IFRVALAGAPVYLMIFDYDTGYTERYMGHPDQNEGYVLGAVAMQAEKFPSEPNRLLLHG 242
DB 730 VFCSGSALSPITDPKLYASAFSERYGLGHGLDNRAVEMAKVAHRVSL--EGQGFVYIHA 787
DB 243 FLDENVHFAHTSILSLFLVAGKRYDQIYPOGHSIRVBSGSHYELHLHYLQE 298
DB 788 TADEKIHFOHTAELITQLINGKANYSLQIYPPDESHYFSSAALQOHLHRSILGFEVE 843

RESULT 12
DPP6_HUMAN STANDARD; PRT; 865 AA.

AC P42658;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE (Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidyl aminopeptidase 6) (Dipeptidyl aminopeptidase IV like protein)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN DPP6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
RC TISSUE=Hippocampus;
RX MEDLINE=93372805; Pubmed=8103397;
RT Yokotani N., Doi K., Wenthold R.J., Wada K.,
RT "Non-conservation of a catalytic residue in a dipeptidyl
RT aminopeptidase IV-related protein encoded by a gene on human
RT chromosome 7.";
RL Hum. Mol. Genet. 2:1037-1039(1993).
CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=DPPX-L;
CC IsoId=P42658-1; Sequence=Displayed;
CC Name=DPPX-S;
CC IsoId=P42658-2; Sequence=VSP_005365;
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC -1- SIMILARITY: Belongs to peptidase family 59B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; M96859; AAA35760.1; -
DR EMBL; M96860; AAA35761.1; -
DR PIR; I54331; I54331.
DR MEROS; S09_973; -
DR Genew; HGNC:3010; DPP6.
DR MIM; 126141; -
DR GO; GO:0008239; F:dipeptidyl-peptidase activity; TAS.
DR InterPro; IPR002469; DPIP N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estr_site.
DR Pfam; PF00930; DPIP N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KM Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 81 MASLYORFTGKINTSRSPAPAPASHLLGCGEEDGAGA
KPLGPRAOAAAPRRGGGAGGAGRRPFOYQGRSDGDEE
-> MTTAKESASAGKSVQOQEOE (in isoform
DPPX-S).
FT FT /FTID=VSP_005365.
FT FT
SQ SEQUENCE 865 AA; 97588 MW; 14B1AE0E024464B CRC64;
Query Match 14.8%; Score 248; DB 1; Length 865;
Best Local Similarity 23.0%; Pred. No. 4.1e-14;
Matches 84; Conservative 53; Mismatches 139; Indels 90; Gaps 10;

CC
DB 1 FEETKDSPLEHNLVSVNPGVETRLTDKRGVSHSCCISQHCDFPISYKSNQKPHC 57
DB 502 FLSTEDLPKRRRLYSAN-----TEGNFNRQCLVDNCTYFASFSHSD--- 548
DB 58 VSLYKSSP-----EDDPCTKTEK-----WATILDSAGPLDYPTEPESFESTT 103
DB 549 FFLKCEGPGVAVNTVNTTDDKKMFDLETNEHVKKAIANDROMPKVEYRIE-----D 602
DB 104 GFTLYGMLYKPHDLQCKKYPTVLFYGGPQG----- 136
DB 603 DYVLPQIILKPAFTPTTTHPILLVVDGTPGSGSVAEKEFVSWETWVSSHGAVVVKCDG 662
DB 137 -----IEIDQVEGLQYLASRYDFIDLVDVGHGMSYGGYLSLMA 176
DB 663 RSGSGFGTKLHEVRRRLGLEEKQMEAVRWML--KEQYIDRTRVAVFGDYGGLSTYI 721
DB 177 LMRSD-----IFRVALAGAPVYLMIFDYDTGYTERYMGHPDQNEGYVLGAVAMQAEKFP 232
DB 722 LPAKGNQOTFTGSGSALSPITDFKLYASAFSERYGLGHGLDNRAVEMKVAHRVSL-- 779
DB 233 EPNRLLLHGFLDENVHFAHTSILSLFLVAGKRYDQIYPOGHSIRVBSGSHYELHL 292
DB 780 EEOGFPIIHPTADEKIHFOHTAELITQLINGKANYSLQIYPPDESHYFSSALQOHLHRSI 839
DB 293 LHYLQE 298
DB 840 INFVEVE 845

RESULT 13
YDZF_SCHPO STANDARD; PRT; 853 AA.

AC Q9P769;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative dipeptidyl aminopeptidase C14C4.15c (EC 3.4.14.-).

GN SPAC14C4.15C OR SPAPU760.01C.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21648401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
 RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
 RA Wellens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Fuchs M., Fritze C., Holzer E., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wandt R., Purnelle B.,
 RA Goffeau A., Godin E., Dietzo S., Gloux S., Lelaur V., Mottier S.,
 RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucena M., Rochet M., Gallardin C., Tallada V., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
 RA Shporkovskii G.V., Usero D., Barrett B.G., Nüsse P.,
 RA "The genome sequence of Schizosaccharomyces pombe".
 RT Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
 CC VACUOLUS.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL162631; CAB83084.1; -
 CC EMBL: Z98596; CAB11208.1; -
 CC GeneDB: Spombe; SPAC14C4.15C; -
 CC InterPro: IPR002469; DPPIV_N term.
 CC InterPro: IPR001375; Peptidase_S9.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC Pfam: PF00930; DPPIV_N term; 1.
 CC Pfam: PF00326; Peptidase_S9; 1.
 CC Hypothetical protein: Hydrolase; Aminopeptidase; Dipeptidase;
 CC Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
 KW DOMAIN 1
 FT TRANSMEM 67 89
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL (POTENTIAL).
 FT ACT_SITE 719 853
 FT ACT_SITE 795 795
 FT ACT_SITE 828 828
 FT ACT_SITE 96 96
 FT CARBOHYD 102 102
 FT CARBOHYD 472 472
 FT CARBOHYD 483 483
 FT CARBOHYD 613 613
 FT CARBOHYD 853 AA; 98341 MW; 384508A50F8304B6 CRC64;
 SO SEQUENCE

Best Local Similarity 29.1%; Pred. No. 9, 1e-14;
 Matches 69; Conservative 28; Mismatches 88; Indels 52; Gaps 8;
 QY 93 PEIIFESFETGFTGLYGLYKPHDLOPKKYPVLFYGGPQ-----GQIED-----140
 DB 602 PSVFFVKIKKNTAVIQEIRPNFPRKRYFVFLHYGAPQALVTGKYMIDNELMAS 661
 QY 141 -----DOVEG-----LOYLASRD--FIDLRVGIHMSYG 169
 DB 662 VYNFLVIKVIDIDISVSGQHLFSDSHELIISKWIELLSYVDTPIIDRRVGIWMSFG 721
 QY 170 GYLISLALMQRSDIPFVALAGAPVTLMIYDPDGYTERWGHDPQNGYVLSVAMQAEK 229
 DB 722 GVLTL-KLENDPFIKGVAVAPVTDKRYDAYISENLIGANSKOTTAIYDKTAHYSEN 780
 QY 230 FSEPNRLLLHGFIDENVFPAHTSILSLFVLAG-----KPYDLOIYPOERHSIRVP 282
 DB 781 F-RKLCGLVLVLTGSDNVHIENTMQLTYAMVEKGVYNYYPF---IVPNANHEFSFP 833
 RESULT 14
 DPP6_MOUSE
 ID DPP6_MOUSE STANDARD; PRT; 804 AA.
 AC Q9Z218; Q9QW42; Q9Z219;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 GN (Dipeptidyl aminopeptidase-related protein) (DPPX).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=99030650; PubMed=9811881;
 RA Hough R.B., Langeling A., Bedian V., Lo C., Bucan M.,
 RA "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
 RA like protein 6 and causes dysregulation of kit expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
 CC -1- FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad (by
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF092507; AAC97366.1; -
 CC EMBL: AF092506; AAC97365.1; -
 CC EMBL: AF092505; AAC98381.1; -
 CC MEROPS: S09.973; -
 CC MGD: MGI:94921; DPP6.
 CC InterPro: IPR002469; DPPIV_N term.
 CC InterPro: IPR001375; Peptidase_S9.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC Pfam: PF00930; DPPIV_N term; 1.
 CC Pfam: PF00326; Peptidase_S9; 1.
 CC Transmembrane; Glycoprotein; Signal-anchor.
 KW DOMAIN 1
 FT TRANSMEM 35 55
 FT SIGNAL-ANCHOR (POTENTIAL).
 FT (POTENTIAL).

Query Match

14.5%; Score 244; DB 1; Length 853;

FT DOMAIN 56 804 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 440 440 MISSING (IN REF. 1; AAC98381).
 FT CONFLICT 638 638 O -> P (IN REF. 1; AAC97365).
 FT CONFLICT 638 638
 SO SEQUENCE 804 AA; 91260 MM; 09CFCE7AADA8A7168 CRC64;
 Query Match 14.3%; Score 241; DB 1; Length 804;
 Best Local Similarity 22.6%; Pred. No. 1.5e-13;
 Matches 82; Conservative 52; Mismatches 145; Indels 84; Gaps 9;
 QY 1 FEGTKDSPLEHNLHYVSVNPGEVTR-----LTDGYSHSCCISQHCDFISKYSNOKN 54
 DB 441 FLSTEDLPRLRLHYSAANTVD--DPNRQUSCDLVENCTVVSASFNMDFLLKCEGPGV 498
 QY 55 PHCV-----SLYKLSPEDDPTCKTKEFMATILDSAGPLDYPPEIFSFESTTGT 106
 DB 499 PTVTANTTDKRMFPLEANE-----EVQKAIKDROMKIEYKRIEV-----EDYS 544
 QY 107 LYGMLYKPHDLQPKKYPTVLFYGGPQG----- 136
 DB 545 LPMQILKPAFTFTDTAHYPLLVVDGTPGSQSVTEREVEWETVLVSHGAVVVKCDGRGS 604
 QY 137 -----IEIDQVEGLOYLASRYDFIDLRVGIHGMVSGYLSLMAIQ 179
 DB 605 GFGTKLHEVRRRLGLFEKQDMQEAVRTWL-KEQYIDKTRVAVFGKDYGGYSTYLPLA 663
 QY 180 RSD-----IFRVALAGAPVTLMIFYDTGYTERYMGHPDQEGYGLSVAMQAEKPESEPN 235
 DB 664 KENOGQFTCGSALPTIDFKLYASAFSERLYGLHGLDRAVEMTKLAHRVSAL--EDQ 721
 QY 236 RLLLLHGFLENVHFAHTSILSFLVRAGKPYDLOIYPOERSIRVPESGEYELHLHY 295
 DB 722 QPFIHATADEKIHFOHTAELLITOLIKKANYSLOIYPPESHYFHSVALKOHLSRSIGF 781
 QY 296 LOE 298
 DB 782 FVE 784
 RESULT 15
 DPP6 RAT STANDARD; PRT; 859 AA.
 ID DPP6 RAT
 AC P46101;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
 DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
 DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
 GN DPP6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN 11;
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 members of a dipeptidyl aminopeptidase family.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:197-201 (1992).
 RL -; FUNCTION: May be involved in the physiological processes of brain
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of
 CC activity may be due to the substitution of an aspartate residue
 CC for the serine residue in the proposed catalytic triad.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=DPPX-L;
 CC Name=DPPX-S;
 CC IsoId=P46101-1; Sequence=Displayed;
 CC IsoId=P46101-2; Sequence=VSP 005366;
 CC -1- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
 CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
 CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
 CC -1- SIMILARITY: Belongs to peptidase family S9B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or bend an email to license@isb-sib.ch).
 CC EMBL; M76426; AAC2061.1; -;
 CC EMBL; M76427; AAC2062.1; -;
 CC MEROPS; S09.973; -;
 CC DR InterPro; IPR002469; DPPXV N term.
 CC DR InterPro; IPR001375; Peptidase_S9.
 CC DR InterPro; IPR000379; Ser. esterase site.
 CC DR Pfam; PFO0930; DPPXV N term; 1.
 CC DR Pfam; PFO0326; Peptidase_S9; 1.
 CC KX Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 CC FT DOMAIN 1 89
 CC FT TRANSMEM 90 110
 CC FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC FT (POTENTIAL).
 CC FT DOMAIN 111 859
 CC FT CARBOHYD 167 167 EXTRACELLULAR (POTENTIAL).
 CC FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT VARSPLIC 1 75
 CC FT MASLYORFGKINTSRSPAPASLILGCGPBEAGSKP
 CC FT LCPQAOAVAPRRGAGGRRPFOYQARSQDEED--> MTT
 CC FT AKPSAGSKSVQOQDO (In isoform DPPX-S).
 CC FT /FTId=VSP 005366.
 SQ SEQUENCE 859 AA; 97301 MM; CE26856D26ED126B CRC64;
 Query Match 14.3%; Score 240; DB 1; Length 859;
 Best Local Similarity 22.3%; Pred. No. 2.1e-13;
 Matches 81; Conservative 55; Mismatches 143; Indels 84; Gaps 9;
 QY 1 FEGTKDSPLEHNLHYVSVNPGEVTRLTDRGYSHSCCISQHCDFISKYSNOKNPHCVSL 60
 DB 496 FLSTEDLPRLRLHYSAANTVD--DPNRQUSCDLVENCTVVSASFNMDFLLKCEGPGV 545
 QY 61 YKLSP-----EDDPCTCKTEF-----WATILDSAGPLDYPPEIFSFESTTGT 106
 DB 546 LKCEGQVPTVTANTTDKRMFPLEANEQVOKAIVDRQMPKIEYKRIEV-----EDYS 599
 QY 107 LYGMLYKPHDLQPKKYPTVLFYGGPQG----- 136
 DB 600 LPMQILKPAFTFTDTAHYPLLVVDGTPGSQSVTEREVEWETVLVSHGAVVVKCDGRGS 659
 QY 137 -----IEIDQVEGLOYLASRYDFIDLRVGIHGMVSGYLSLMAIQ 179
 DB 660 GFGTKLHEVRRRLGLFEKQDMQEAVRTWL-KEQYIDKTRVAVFGKDYGGYSTYLPLA 718
 QY 180 RSD-----IFRVALAGAPVTLMIFYDTGYTERYMGHPDQEGYGLSVAMQAEKPESEPN 235
 DB 719 KENOGQFTCGSALPTIDFKLYASAFSERLYGLHGLDRAVEMTKLAHRVSAL--EDQ 776
 QY 236 RLLLLHGFLENVHFAHTSILSFLVRAGKPYDLOIYPOERSIRVPESGEYELHLHY 295

Db	777	QFLIHATADEKIHFOHTAELITOLIKKANVSLQIYPDESHYFHSVALKOHLYRSIIGF	836
Qy	296	LQE	298
Db	837	FVE	839

Search completed: October 15, 2003, 17:11:28
Job time : 6.9182 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 ; Search time 19.8265 Seconds

(without alignments)
4034.822 Million cell updates/sec

Title: US-10-070-464-3
Perfect score: 1680
Sequence: 1 FEGTKDSPLEHLYVSVYN.....HLHLYQENLGRNALKVI 310

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_proteus:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1680	100.0	310	4 Q9HBM4	Q9HBM4 homo sapien
2	1645.5	97.9	882	4 Q9HBM5	Q9HBM5 homo sapien
3	1645.5	97.9	882	4 Q9HBM7	Q9HBM7 homo sapien
4	1605.5	95.6	892	11 Q9D4G6	Q9D4G6 mus musculu
5	1407	83.8	831	4 Q8NEM5	Q8NEM5 mus musculu
6	1401	83.4	632	4 Q96JX1	Q96JX1 homo sapien
7	1377.5	82.0	465	4 Q9HBM3	Q9HBM3 homo sapien
8	1094.5	65.1	862	11 Q9HBM9	Q9HBM9 mus musculu
9	1093.5	65.1	439	4 Q9BVR3	Q9BVR3 homo sapien
10	1093.5	65.1	862	11 Q8BVG4	Q8BVG4 mus musculu
11	1093.5	65.1	863	4 Q8WMD8	Q8WMD8 homo sapien
12	1086.5	64.7	360	4 Q9HBM2	Q9HBM2 homo sapien
13	1032.5	57.6	508	4 Q75273	Q75273 homo sapien
14	968.5	57.6	628	4 Q8N3F5	Q8N3F5 homo sapien
15	938.5	55.9	312	4 Q96WTF	Q96WTF homo sapien
16	708	42.1	1042	5 Q9VC20	Q9VC20 drosophila

17	708	42.1	1102	5 Q9VC19	Q9VC19 drosophila
18	704.5	39.2	1053	5 Q8IH07	Q8IH07 drosophila
19	659	39.2	167	4 Q8N2J7	Q8N2J7 homo sapien
20	502	29.9	927	5 Q965K3	Q965K3 caenorhabdi
21	502	29.9	931	5 Q44987	Q44987 caenorhabdi
22	484.5	28.8	738	16 Q9A6E0	Q9A6E0 caulobacter
23	442	26.3	552	10 Q8GJY7	Q8GJY7 arabidopsis
24	442	26.3	766	10 Q9FNF6	Q9FNF6 arabidopsis
25	441.5	26.3	757	16 Q8FED7	Q8FED7 xanthomonas
26	436.5	26.0	753	16 Q8EAB7	Q8EAB7 shewanella
27	433.5	25.8	751	16 Q8P3V8	Q8P3V8 xanthomonas
28	398.5	23.7	711	2 P95782	P95782 xanthomonas
29	379	22.6	711	2 Q47900	Q47900 flavobacter
30	375	22.4	723	2 Q31048	Q31048 porphyromon
31	375	22.4	723	2 Q66223	Q66223 porphyromon
32	374.5	22.3	730	2 Q93JY4	Q93JY4 prevotella
33	343.5	20.4	901	3 Q96VT7	Q96VT7 aspergillus
34	341	20.3	748	13 P70092	P70092 xenopus lae
35	334	19.9	237	2 Q87543	Q87543 capnocytoph
36	328	19.5	802	5 Q9VTH1	Q9VTH1 drosophila
37	327	19.5	745	5 Q9VTH2	Q9VTH2 drosophila
38	327	19.5	771	3 Q42812	Q42812 aspergillus
39	315.5	18.8	765	3 Q14425	Q14425 aspergillus
40	314.5	18.7	761	11 Q8R492	Q8R492 rattus norv
41	305.5	18.2	935	5 Q9VMB4	Q9VMB4 drosophila
42	304.5	18.1	827	16 Q8PPU4	Q8PPU4 xanthomonas
43	303.5	18.1	799	5 Q18119	Q18119 caenorhabdi
44	293	17.4	707	16 Q9F348	Q9F348 streptomyce
45	284.5	16.9	829	5 Q18253	Q18253 caenorhabdi

ALIGNMENTS

RESULT 1	Q9HBM4	PRELIMINARY;	PRT;	310 AA.
ID	Q9HBM4			
AC	Q9HBM4			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Dipeptidyl peptidase 8 (Fragment).			
GN	DPPE8.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=20467194; PubMed=11012666;			
RA	Abdott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,			
RA	Goirell M.D.,			
RT	"Cloning, expression and chromosomal localization of a novel human			
RT	dipeptidyl peptidase (DPP) IV homolog, DPP8.";			
RL	Eur J Biochem. 267:6140-6150(2000).			
DR	EMBL; AF221635; AAC29767.1; ..			
DR	MEDLINE; S09.018; ..			
DR	InterPro; IPR001375; Peptidase_S9.			
DR	Pfam; PF00326; Peptidase_S9; 1.			
FT	NON_TER			
FT	SEQUENCE 310 AA; 35396 MW; 8E87C3402ED9C7AC CRC64;			
QY	Query Match	100.0%; Score 1680; DB 4; Length 310;		
QY	Best Local Similarity	100.0%; Pred. No. 9.7e-145;		
QY	Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
DB	1 FEGTKDSPLEHLYVSVYNPEGEVRLTDGSGSCCTSOHCDPFRISKSNQKPRCVSL 60			
DB	1 FEGTKDSPLEHLYVSVYNPEGEVRLTDGSGSCCTSOHCDPFRISKSNQKPRCVSL 60			
QY	61 YKLSPEDDPTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYGMLKKPHDLP 120			

```

Db      61 YKLSDEDDPTCKTKEFWATILDSAGPLPDYTPPELIFSFEISTGFTLYGMLYKPHDLOPG 120
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      181 SDIFRVAIAGAPYTLWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLL 240
      181 SDIFRVAIAGAPYTLWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLL 240
      181 SDIFRVAIAGAPYTLWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLL 240
      181 SDIFRVAIAGAPYTLWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLL 240
      241 HGLDENVHFAHNSILSLFVRACKPDLQIYPOERHSIRVPESGEHYELHLHYOENL 300
      241 HGLDENVHFAHNSILSLFVRACKPDLQIYPOERHSIRVPESGEHYELHLHYOENL 300
      241 HGLDENVHFAHNSILSLFVRACKPDLQIYPOERHSIRVPESGEHYELHLHYOENL 300
      301 GSRIALKVI 310
      301 GSRIALKVI 310
      301 GSRIALKVI 310

```

RESULT 2

```

ID      09HBM5      PRELIMINARY:      PRT:      882 AA.
AC      09HBM5;
DT      01-MAR-2001 (T-EMBLrel. 16, Created)
DT      01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT      01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE      Dipeptidyl peptidase 8.
GN      DPP8.
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RX      MEDLINE=20467194; Pubmed=11012666;
RA      Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA      Gorell M.D.;
RT      "Cloning, expression and chromosomal localization of a novel human
RT      dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL      Eur. J. Biochem. 267:6140-6150(2000).
DR      EMBL: AF221634; AAG29766.1;
DR      MEROPS: S09_018;
DR      InterPro: IPR002469; DPP1V_N term.
DR      InterPro: IPR001375; Peptidase_S9.
DR      InterPro: IPR000379; Ser_estra_site.
DR      Pfam: PF00930; DPP1V_N term; 1.
DR      Pfam: PF00326; Peptidase_S9; 1.
SQ      SEQUENCE 882 AA; 101421 MW; AD801C302DB4652B CRC64;

```

```

Query Match      97.9%; Score 1645.5; DB 4; Length 882;
Best Local Similarity 86.4%; Pred. No. 5.5e-141;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```

```

QY      1 FEETKOSPLEHNLVYVSYNVPGEVTRLTDGYSHSCISQHCDFPISKYSNQKPHCVSL 60
      524 FEETKOSPLEHNLVYVSYNVPGEVTRLTDGYSHSCISQHCDFPISKYSNQKPHCVSL 583.
      61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPELIFSFEISTGFTLYGMLYKPHDLOPG 120
      584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPELIFSFEISTGFTLYGMLYKPHDLOPG 643
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      644 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMORSDIFRVAIAGA 191
      704 YKMGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMORSDIFRVAIAGA 763
      192 PVTLMWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLDENVHFA 251
      764 PVTLMWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLDENVHFA 823

```

```

QY      252 HTSILSLFVRACKPYDLQIYPOERHSIRVPESGEHYELHLHYOENLGSRIALKVI 310
      824 HTSILSLFVRACKPYDLQIYPOERHSIRVPESGEHYELHLHYOENLGSRIALKVI 882

```

RESULT 3

```

ID      081MG7      PRELIMINARY:      PRT:      882 AA.
AC      081MG7;
DT      01-MAR-2003 (T-EMBLrel. 23, Created)
DT      01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE      Similar to dipeptidyl peptidase 8.
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RX      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Strausberg R.;
RL      Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL: BE040204; AAA40203.1;
SQ      SEQUENCE 882 AA; 101391 MW; 88C76AF5BCE707F9 CRC64;

```

```

Query Match      97.9%; Score 1645.5; DB 4; Length 882;
Best Local Similarity 86.4%; Pred. No. 5.5e-141;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```

```

QY      1 FEETKOSPLEHNLVYVSYNVPGEVTRLTDGYSHSCISQHCDFPISKYSNQKPHCVSL 60
      524 FEETKOSPLEHNLVYVSYNVPGEVTRLTDGYSHSCISQHCDFPISKYSNQKPHCVSL 583
      61 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPELIFSFEISTGFTLYGMLYKPHDLOPG 120
      584 YKLSPEDDPTCKTKEFWATILDSAGPLPDYTPPELIFSFEISTGFTLYGMLYKPHDLOPG 643
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      121 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      644 KKPYTULFIYGGOGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMOR 180
      135 ---GQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMORSDIFRVAIAGA 191
      704 YKMGQIEIDDOVEGLOYLASRYDFIDLDRVGIHGSYGYLSLMMORSDIFRVAIAGA 763
      192 PVTLMWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLDENVHFA 251
      764 PVTLMWIFDYDGYTERYMGHPDQNEQGYLGSVAMQAEKFPSEPNRLLLHGLDENVHFA 823
      252 HTSILSLFVRACKPYDLQIYPOERHSIRVPESGEHYELHLHYOENLGSRIALKVI 310
      824 HTSILSLFVRACKPYDLQIYPOERHSIRVPESGEHYELHLHYOENLGSRIALKVI 882

```

RESULT 4

```

ID      09D4G6      PRELIMINARY:      PRT:      892 AA.
AC      09D4G6;
DT      01-JUN-2001 (T-EMBLrel. 17, Created)
DT      01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE      Adult male testis cDNA, RIKEN full-length enriched library,
DE      clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
OS      Mus musculus (mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      (1)
RX      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=Testis;
RA      Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA      Arawaka T., Bono H., Carninci P., Fukuda S., Fukushima Y., Furuno M.,
RA      Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hoti F.,

```

RA Imotani K., Iehli Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RA Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Mech. Enzymol. 303:19-44(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama Y., Nishi K., Kitsuai T., Ishihiro H., Itoh M.,
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-Format
RT sequencing pipeline with 384 multichipillary sequencer."
RL Genome Res. 10:1757-1771(2000).
DR EMBL: AK016546; BAB30295.2; -
SQ SEQUENCE 892 AA; 102284 MW; FDE0DBEDCA4CA346 CRC64;

Query Match 95.6%; Score 1605.5; DB 11; Length 892;
Best Local Similarity 84.1%; Pred. No. 2.5e-137;
Matches 302; Conservative 3; Mismatches 5; Indels 49; Gaps 1;

QY 1 FEGTKDSPLEHLLVVSIVNPEVRLTDRGYSGCCISOHCDPFIKSYNOKNPHCVSL 60
DB 534 FEGTKDSPLEHLLVVSIVNPEVRLTDRGYSGCCISOHCDPFIKSYNOKNPHCVSL 593
QY 61 YKLSPPEDPTCKTEKFPATILDSAGPLDYTPPEIFSFESTTGTTLYGMLYKPHDLOG 120
DB 594 YKLSPPEDPTCKTEKFPATILDSAGPLDYTPPEIFSFESTTGTTLYGMLYKPHDLOG 653
QY 121 KKYPTVLFYGGPO----- 134
DB 654 KKYPTVLFYGGPOVQLVNNRFGVKYFRLNTLASLGVVVVVINDRGSGRGLKFECAF 713

QY 135 ---GGIEIDPOVEGIQYLASRYDFIDLRVGHGSSYGYLSIMLMQSRDIFRVAIAGA 191
DB 714 YKMGIEIDPOVEGIQYLASRYDFIDLRVGHGSSYGYLSIMLMQSRDIFRVAIAGA 773
QY 192 PVTLLMIFYDTGTERMYGHPDNEGGYLYGSVAMQAEKFPSEPNRLLLHGLDENVHFA 251
DB 774 PVTLLMIFYDTGTERMYGHPDNEGGYLYGSVAMQAEKFPSEPNRLLLHGLDENVHFA 833
QY 252 HTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHLHYQENIGSRILAKVI 310
DB 834 HTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHLHYQENIGSRILAKVI 892

RESULT 5

Q8NEM5 PRELIMINARY; PRT; 831 AA.
AC Q8NEM5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Similar to dipeptidylpeptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Streuberg R.;
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC030688; AAH30688.1; -
SQ SEQUENCE 831 AA; 93527 MW; 0B2A13A2FE70CBE2 CRC64;

Query Match 83.8%; Score 1407; DB 4; Length 831;
Best Local Similarity 84.9%; Pred. No. 2.7e-119;
Matches 270; Conservative 11; Mismatches 19; Indels 18; Gaps 5;

QY 1 FEGTKDSPLEHLLVVSIVNPEVRLTDRGYSGCCISOHCDPFIKSYNOKNPHCVSL 60
DB 524 FEGTKDSPLEHLLVVSIVNPEVRLTDRGYSGCCISOHCDPFIKSYNOKNPHCVSL 583
QY 61 YKLSPPEDPTCKTEKFPATILDSAGPLDYTPPEIFSFESTTGTTLYGMLYKPHDLOG 120
DB 584 YKLSPPEDPTCKTEKFPATILDSAGPLDYTPPEIFSFESTTGTTLYGMLYKPHDLOG 643
QY 121 KKYPTVLFYGGPOQIIEIDPOVEGLQY----LAS-RYDFIDLRVGI--HGWSYGYL 172
DB 644 KKYPTVLFYGGPOVQLVNNRFGVKYFRLNTLASLGVVVVVINDRGSGRGLKFECAF 702
QY 173 SIMLMQSRDIFRVAIAGAPVTLMIFYDTGTERMYGHPDNEGGYLYGSVAMQAEKPPS 232
DB 703 KYK-----MVAIAGAPVTLMIFYDTGTERMYGHPDNEGGYLYGSVAMQAEKPPS 753
QY 233 EPNRLLLHGLDENVHFAHTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHL 292
DB 754 EPNRLLLHGLDENVHFAHTSILSLVLRAGKPYDQIYPOERHSIRVPESGSEHYELHL 813
QY 293 LHYLQENIGSRILAKVI 310
DB 814 LHYLQENIGSRILAKVI 831

RESULT 6

Q96JX1 PRELIMINARY; PRT; 632 AA.
AC Q96JX1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein Fluj14920.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Makatsuta A., Nakamura Y., Nagahara K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027826; BAB55395.1; -.
 DR MEROPS; S09.018; -.
 KW Hypothetical protein.
 SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;
 Query Match 83.4%; Score 1401; DB 4; Length 632;
 Best Local Similarity 84.6%; Pred. No. 6,5e-119;
 Matches 269; Conservative 11; Mismatches 20; Indels 18; Gaps 5;
 QY 1 FEGTDSPLEHLLVYVSYVNPGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 335 FEGTDSPLEHLLVYVSYVNPGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 384
 QY 61 YLSSPEDPTCKTEFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOG 120
 DB 365 YLSSPEDPTCKTEFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOG 444
 QY 121 KKYPTVLFIYGGPQOIEIDOVEGLQY-----LAS-RYDIDLDRVGI--HGWSYGYL 172
 DB 445 KKYPTVLFIYGGPQOYOL--VNNRFKGVKPYRLNTLASLGVVVVIDNRGSCHEGLKEGAF 503
 QY 173 SLMALMQRSDIFRVNAGAPVTLMIFYDTGYTERVYGHDPQNEQGYLGSAVMAQAEKFP 232
 DB 504 KYK-----MVALAGAPVTLMIYFDGYTERVYGHDPQNEQGYLGSAVMAQAEKFP 554
 QY 233 EBNRLLLHGFLENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSIRVPSGEHVELH 292
 DB 555 EBNRLLLHGFLENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSIRVPSGEHVELH 614
 QY 293 LHYLOENLGSRIALAKVI 310
 DB 615 LHYLOENLGSRIALAKVI 632
 RESULT 7
 Q9HBM3 PRELIMINARY; PRT; 465 AA.
 AC Q9HBM3;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Dipeptidyl peptidase 8 (Fragment).
 GN DPP8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RC MEDLINE=20467194; PubMed=11012666;
 RA Abbott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,
 RA Gorell M.D.,
 RT "Cloning, expression and chromosomal localization of a novel human
 RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
 RL Eur. J. Biochem. 267:6140-6150(2000).
 DR EMBL; AF221636; AAG29768.1; -.
 DR MEROPS; S09.018; -.
 FT NON_TER 1 1

SQ SEQUENCE 465 AA; 53197 MW; 22939EC0A4EE29CE CRC64;
 Query Match 82.0%; Score 1377.5; DB 4; Length 465;
 Best Local Similarity 83.5%; Pred. No. 5,9e-117;
 Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;
 QY 1 FEGTDSPLEHLLVYVSYVNPGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 207 FEGTDSPLEHLLVYVSYVNPGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 266
 QY 61 YLSSPEDPTCKTEFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOG 120
 DB 267 YLSSPEDPTCKTEFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOG 326
 QY 121 KKYPTVLFIYGGPQOIEIDOVEGLQYLASRYDFIDLDRVGIHGWSYGYLSLMAIMQR 180
 DB 327 KKYPTVLFIYGGPQ----- 340
 QY 181 SDIFRVNAGAPVTLMIFYDTGYTERVYGHDPQNEQGYLGSAVMAQAEKFPENRLL 240
 DB 341 -----VALAGAPVTLMIFYDTGYTERVYGHDPQNEQGYLGSAVMAQAEKFPENRLL 395
 QY 241 HGFLENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSIRVPSGEHVELHLYLOENL 300
 DB 336 HGFLENVHFAHTSILLSFLVRAGKPYDLOIYPOERHSIRVPSGEHVELHLYLOENL 455
 QY 301 GSRIALAKVI 310
 DB 456 GSRIALAKVI 465
 RESULT 8
 Q9BWT9 PRELIMINARY; PRT; 862 AA.
 AC Q9BWT9;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Dipeptidyl peptidase 9 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RC MEDLINE=22354683; PubMed=12466851;
 RA The KIKEN Genome Exploration Research Group Phase I & II Team,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK050021; BAC34034.1; -.
 SQ SEQUENCE 862 AA; 98001 MW; BID566E824A834E8 CRC64;
 Query Match 65.1%; Score 1094.5; DB 11; Length 862;
 Best Local Similarity 57.3%; Pred. No. 8,1e-91;
 Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;
 QY 1 FEGTDSPLEHLLVYVSYVNPGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 60
 DB 514 FEGTDSPLEHLLVYVSYVNPGEVTRLTDGYSHSCCISQHCDFISKYSNQNPHCVSL 573
 QY 64 YLSSPEDPTCKTEFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOG 120
 DB 574 YLSSPEDPTCKTEFMAITLDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLOG 633
 QY 121 KKYPTVLFIYGGPQ----- 134
 DB 634 KKYPTVLFIYGGPQOYOLVNNRFKGVKPYRLNTLASLGVVVVIDNRGSCHEGLKEGAF 693
 QY 135 ---GOIEIDOVEGLQYLASRYDFIDLDRVGIHGWSYGYLSLMAIMQRSDIFRVNAG 191

Db 694 NMGQVEIEDQVEGLQVAAEKYGFIDLSRAVAIHGWSYGFSLMGLIHKEQVFAIAGA 753
 QY 192 PVTLMIFYDTGYTERYWGHPDQNEQGYLGSVAMQAEKFPSEPRRLILHGFIDENVHFA 251
 Db 754 PVTVMAYDGTGYTERYMDVBNNOQGYEASVALHVEKLPNEPRRLILHGFIDENVHFF 813
 QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 300
 Db 814 HTMFVLVSQILRAGKPYQLOIYPERHSIRCESEGEHYEVLLHFLQENHL 862

RESULT 9

Q9BVR3 PRELIMINARY; PRT; 439 AA.
 AC Q9BVR3;
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP TISSUE=Placenta;
 RA Strausberg R.;
 DR EMBL:BC00970; AAH00970.1; --
 DR MEROPS; S09.019; --
 DR GeneW; HGNC:18648; DPP9.
 DR InterPro; IPR001375; Peptidase_S9.
 DR InterPro; IPR000379; Ser esterase.
 DR Pfam; PF00326; Peptidase_S9; 1.
 KM Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 439 AA; 49926 MW; A18BBA9E12092BAF CRC64;

Query Match 65.1%; Score 1093.5; DB 4; Length 439;
 Best Local Similarity 57.3%; Pred. No. 4e-91;

Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 1 FEGTKDSPLEHNLVYVSYNPGVETRLTDGYSHSCISQHCDFIISKYSNQNPHCVSL 60
 Db 91 FQGTCKDTPLEHNLVYVSYNPGVETRLTDGYSHSCISQHCDFIISKYSNQNPHCVSL 150
 QY 61 YKLSSEDDPTCKTKERMAITLDSAGPLPYTPPEIFSESTTGFTLYGMLYKPHDLP 120
 Db 151 YKLSGPDDEPLHQPFRWASMEANCPDPYVPEIFHFTRADVQVLYGMIYKPHDLP 210
 QY 121 KKYPTVLFYGGPQ----- 134
 Db 211 KKHPTVLFYGGPQV 270
 QY 135 ---GQIEIDQVEGLQVLAARYDFIDLDRVGIHGWSYGYLSMALMQRSDFRVAIAGA 191
 Db 271 NMGQVEIEDQVEGLQVAAEKYGFIDLSRAVAIHGWSYGFSLMGLIHKEQVFAIAGA 330
 QY 192 PVTLMIFYDTGYTERYWGHPDQNEQGYLGSVAMQAEKFPSEPRRLILHGFIDENVHFA 251
 Db 331 PVTVMAYDGTGYTERYMDVBNNOQGYEASVALHVEKLPNEPRRLILHGFIDENVHFF 390
 QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 300
 Db 391 HTMFVLVSQILRAGKPYQLOIYPERHSIRCESEGEHYEVLLHFLQENHL 439

RESULT 10

Q8BVG4 PRELIMINARY; PRT; 862 AA.
 AC Q8BVG4;
 DT 01-MAR-2003 (TEMBLrel. 23, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Dipeptidyl peptidase 9 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP STRAIN=CS7BL/6J; TISSUE=Olfactory brain;
 RX MEDLINE=22354683; PubMed=1246851;
 RA The FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL:AK078301; BAC37211.1; --
 SQ SEQUENCE 862 AA; 98049 MW; B4FD3476B6F50030 CRC64;

Query Match 65.1%; Score 1093.5; DB 11; Length 862;
 Best Local Similarity 57.3%; Pred. No. 1e-90;

Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 1 FEGTKDSPLEHNLVYVSYNPGVETRLTDGYSHSCISQHCDFIISKYSNQNPHCVSL 60
 Db 514 FQGTCKDTPLEHNLVYVSYNPGVETRLTDGYSHSCISQHCDFIISKYSNQNPHCVSL 573
 QY 61 YKLSSEDDPTCKTKERMAITLDSAGPLPYTPPEIFSESTTGFTLYGMLYKPHDLP 120
 Db 574 YKLSGPDDEPLHQPFRWASMEANCPDPYVPEIFHFTRADVQVLYGMIYKPHDLP 633
 QY 121 KKYPTVLFYGGPQ----- 134
 Db 634 KKHPTVLFYGGPQVQ 693
 QY 135 ---GQIEIDQVEGLQVLAARYDFIDLDRVGIHGWSYGYLSMALMQRSDFRVAIAGA 191
 Db 694 NMGQVEIEDQVEGLQVAAEKYGFIDLSRAVAIHGWSYGFSLMGLIHKEQVFAIAGA 753
 QY 192 PVTLMIFYDTGYTERYWGHPDQNEQGYLGSVAMQAEKFPSEPRRLILHGFIDENVHFA 251
 Db 754 PVTVMAYDGTGYTERYMDVBNNOQGYEASVALHVEKLPNEPRRLILHGFIDENVHFF 813
 QY 252 HTSILSFLVRAGKPYDLOIYPOERHSIRVSESGEYELHLHYLOENTL 300
 Db 814 HTMFVLVSQILRAGKPYQLOIYPERHSIRCESEGEHYEVLLHFLQENHL 862

RESULT 11

Q8MXD8 PRELIMINARY; PRT; 863 AA.

AC Q8MXD8;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Dipeptidyl peptidase 9.
 GN DPP9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP Olsen C.; Wagmann N.;
 RA "Identification and characterization of a novel member of the
 RT dipeptidyl peptidase IV-related family."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN (2)

RP TISSUE=Skin;
 RC Strausberg R.;
 DR EMBL:AF452102; AAL47179.1; --
 EMBL:BC037948; AAH37948.1; --


```

OY 252 HTSILSLFLVAGKPYDL-----QIYQERHSIRVPSGEGHYEHLHLHYLOENTL 300
DB 451 HTNFLVSQLIRAGKRYQLQVALPVPSPQIYPERHSIRCPESGEGHYEVLTLHLFLOEYL 508

RESULT 14
OBN3F5 PRELIMINARY; PRT; 628 AA.
ID OBN3F5;
AC OBN3F5;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP762M2413.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Ansoore W., Warkner U., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL884376; CAD39039.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00326; Peptidase_S9; 1.
DR Hypothetical protein.
FT NON TER
SQ SEQUENCE 628 AA; 71368 MW; DC0B6A9440507C3B CRC64;

Query Match 57.6%; Score 968.5; DB 4; Length 628;
Best Local Similarity 55.1%; Pred. No. 1.6e-79;
Matches 177; Conservative 40; Mismatches 55; Indels 49; Gaps 1;

OY 1 PEGTKDSLEHLHLVYSVNPGVETRLTDRGYSHSCCISQCHDFITKYSNOKNPHCVSL 60
DB 307 FQGRKDPLEHLHYVSYEAAGEIYRLTTPGSHSCSOSNPFDMVSHYSSVSTPPCVH 366
OY 61 YKSSPEDPTCKTEKFPATILDSAGPLPDYTPRETFESGRTGLYGMLYKPHDLOPG 120
DB 367 YKLSGPDDDLHKOPRFWASMEASCPDYVPPPIFHFTRSDVRLYGMLYKPHALOPG 426
OY 121 KKYPPTVLFLYGGPQ----- 134
DB 427 KKHPTVLFLVYGPGVQLVNNSEFKGIKYLRLNTLASLGAVVVIDGRSGCQGLRFEGLK 486
OY 135 ---GQIETDDQVEGLQYIASRYDFIDLRVGIHGSYSGYLSLALMORSIFRYAIAGA 191
DB 487 NQMGQVEIEDQVEGLQYIAEKYGFIDLSRVAIHGMSYSGYLSLGLIHKPQVFAIAGA 546
OY 192 PVTLMIFDYDGYTRSRWGHDPONEGQYVLGSVAMQAEKFPSEPRILLHLGFLDENYFA 251
DB 547 PVTVMAYDGYTRSRWGHDPONEGQYVAGSVALHVEKLPNEPNRLTLHLGFLDENYHF 606
OY 252 HTSILSLFLVAGKPYDLQIY 272
DB 607 HTNFLVSQLIRAGKRYQLQY 627

RESULT 15
O96NT8 PRELIMINARY; PRT; 312 AA.
ID O96NT8;
AC O96NT8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical protein FLJ30094.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Osuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Maeno Y., Nagai K., Isegai T.,
RT "NEBO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054656; BAB70784.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser esterase site.
DR Pfam; PF00326; Peptidase_S9; 1.
DR Hypothetical protein.
SQ SEQUENCE 312 AA; 35518 MW; ABE940AFC5877717 CRC64;

Query Match 55.9%; Score 938.5; DB 4; Length 312;
Best Local Similarity 55.4%; Pred. No. 3.3e-77;
Matches 173; Conservative 38; Mismatches 52; Indels 49; Gaps 1;

OY 38 ISQCHDFITKYSNOKNPHCVSLYKSSPEDPTCKTEKFPATILDSAGPLPDYTPPEIF 97
DB 1 MSQNFDMVSHYSSVSTPPCVHVVLSGPDDDLHKOPRFWASMEASCPDYVPPPIF 60
OY 98 SPESRTGFTLYGMLYKPHDLOPGKKYPTVFLFYGGPQ----- 134
DB 61 HFTRSDVRLYGMLYKPHALOPGKHPVTLFLVYGPGVQLVNNSEFKGIKYLRLNTLASLG 120
OY 135 -----GQIETDDQVEGLQYIASRYDFIDLRVGIHGSY 168
DB 121 YAVVVIDGRSGCQGLRFEGLKXNMGQVEIEDQVEGLQYIAEKYGFIDLSRVAIHGMSY 180
OY 169 GGYLSLALMORSIFRYAIAGAPVTLMTFYDGYTRSRWGHDPONEGQYVLGSVAMQAE 228
DB 181 GGFLSLMGLIHKPQVFAIACAPVTVMAYDGYTRSRWGHDPONEGQYVAGSVALHVE 240
OY 229 KFPSEPRILLHLGFLDENYFAHTSILSLFLVAGKRYDLQIYQERHSIRVPSGEGHY 288
DB 241 KLPNEPNRLTLHLGFLDENYHFHTNFLVSQLIRAGKRYQLQIYPERHSIRCPESGHY 300
OY 289 ELHLHLHYLOENTL 300
DB 301 EVTLHLFLOEYL 312

```

Search completed: October 15, 2003, 17:13:43
 Job time : 21.8265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:15:46 (Search time 3384.49 Seconds

(without alignment)
3747.092 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680

Sequence: 1 FEGTKDSPLEHLYVSVVN.....HLHLVLENLGRIRALKVI 310

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10070464/runat_15102003_113553_24810/app_query.fasta_1.2652
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464 @CGN 1.1 13561 @runat 15102003 113553 24810 -NCPU=6 -ICPU=3
-NO MAP -LANG=JUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -MAPN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vi: *
30: em_hg_hum: *
31: em_hg_inv: *
32: em_hg_other: *
33: em_hg_mus: *
34: em_hg_pln: *
35: em_hg_rod: *
36: em_hg_mam: *
37: em_hg_vrt: *
38: em_sy: *
39: em_hg_hum: *
40: em_hg_mus: *
41: em_hggo_other: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1680	100.0	1197	9 AF221635	AF221635 Homo sapi
2	1645.5	97.9	2649	6 AX354795	AX354795 Sequence
3	1645.5	97.9	2656	9 AY172659	AY172659 Homo sapi
4	1645.5	97.9	2671	6 AX608725	AX608725 Sequence
5	1645.5	97.9	3106	6 AX342633	AX342633 Sequence
6	1645.5	97.9	3127	9 AF221634	AF221634 Homo sapi
7	1645.5	97.9	3143	6 AX354793	AX354793 Sequence
8	1645.5	97.9	4535	6 BC040203	BC040203 Homo sapi
9	1644.5	97.9	4676	6 AX608743	AX608743 Sequence
10	1625.5	96.8	4309	6 AX608737	AX608737 Sequence
11	1625.5	96.8	4829	6 AX608735	AX608735 Sequence
12	1605.5	95.6	4799	10 BC043124	BC043124 Mus muscu
13	1407	83.8	3130	9 BC030688	BC030688 Homo sapi
14	1401	83.4	2161	6 BD157001	BD157001 Primer fo
15	1401	83.4	2161	9 AK027826	AK027826 Homo sapi
16	1377.5	82.0	1659	9 AF221636	AF221636 Homo sapi
17	1377.5	82.0	2668	6 AX405771	AX405771 Sequence
18	1377.5	82.0	2778	6 AK000290	AK000290 Homo sapi
19	1377.5	82.0	2842	6 AX405770	AX405770 Sequence
20	1372	81.7	2510	6 AX338497	AX338497 Sequence
21	1350.5	80.9	4685	6 AX608745	AX608745 Sequence
22	1352	80.5	4523	6 AX608731	AX608731 Sequence
23	1093.5	65.1	2546	9 BC000970	BC000970 Homo sapi
24	1093.5	65.1	2602	9 AY172660	AY172660 Homo sapi
25	1093.5	65.1	2617	6 AX608727	AX608727 Sequence
26	1093.5	65.1	2676	6 AX524928	AX524928 Sequence
27	1093.5	65.1	3000	9 AF542510	AF542510 Homo sapi
28	1093.5	65.1	3716	6 AX480934	AX480934 Sequence
29	1093.5	65.1	4219	6 AX608751	AX608751 Sequence
30	1093.5	65.1	4232	9 BC037948	BC037948 Homo sapi
31	1093.5	65.1	4295	9 AF452102	AF452102 Homo sapi
32	1093.5	65.1	4302	6 AX608747	AX608747 Sequence
33	1086.5	64.7	1083	9 AF221637	AF221637 Homo sapi
34	1086.5	64.7	4180	6 AX608759	AX608759 Sequence
35	1086.5	64.7	4263	6 AX608757	AX608757 Sequence
36	1084.5	64.6	2261	6 AX713363	AX713363 Sequence
37	1084.5	64.6	2261	9 AK054656	AK054656 Homo sapi
38	968.5	57.6	3243	9 HSM805448	AK054656 Homo sapi
39	934	55.6	873	6 AX081330	AL813376 Homo sapi
40	929	55.3	735	6 AX524942	AX081330 Sequence
41	929	55.3	1002	6 AX524939	AX524942 Sequence
42	872.5	51.9	4076	6 AX608755	AX524939 Sequence
43	872.5	51.9	4159	6 AX608753	AX608755 Sequence
44	865.5	51.5	4037	6 AX608763	AX608753 Sequence
45	865.5	51.5	4120	6 AX608761	AX608763 Sequence

RESULT 1

ALIGNMENTS

AF221635
 LOCUS AF221635 1197 bp mRNA linear PRI 05-NOV-2000
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds, alternatively spliced.
 ACCESSION AF221635
 VERSION AF221635.1 GI:11095189
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1197)
 Abbott,C.A., Yu,D.W., Woollett,E., Sutherland,G.R., McCaughan,G.W. and Gorrell,M.D.
 Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8
 Eur. J. Biochem. 267 (20), 6140-6150 (2000)
 JOURNAL MEDLINE
 PUBMED 11012666
 REFERENCE 2 (bases 1 to 1197)
 Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.
 Direct Submission
 Submitted (06-JUN-2000) A.W. Morrow Gastroenterology and Liver Centre, Centenary Institute of Cell Biology and Cancer Medicine, Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia
 JOURNAL
 TITLE Location/Qualifiers
 FEATURES
 source 1. .1197
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q22"
 /clone="T21"
 /issue_type="Placenta"
 /clone_lib="lambda triplexXTM: Clontech"
 <1..1197
 /gene="DPP8"
 CDS
 <1..935
 /gene="DPP8"
 /note="serine protease; member of prolyl oligopeptidase S9b family belonging to enzyme clan SC; additional splice variants of this mRNA are found in GenBank Accession Numbers AF221634, AF221636 and AF221637"
 /codon_start=3
 /product="dipeptidyl peptidase 8"
 /protein_id="AAC29767.1"
 /db_xref="GI:11095190"
 /translation="FEGTKDSPLEEHLYVSVYVNGEVTRLTDPGVSQSCISQCHP FISKYSQKRNPHCVSLKLSPEDDPTCKTEFNATTLDSAGPLPDYTPPEISFEST TGFLLYGLMLYKPHDLPQPKYPTVLFYGGQSGIETDDVEGLQYIASRYDFEST VGIHGSVYGGYGLSLMALMQRSDIFRVAIAGAPVLTWTFDYGYTERVWGHDPDQEGY YLGSVAQAQAEKFESEPNRLILHGFLENFAHTSILLSFLVAGKRPYDQIVPQGR HSIRVPSGEHYELHLHYLOENLGSRIALAKVI"
 BASE COUNT 357 a 258 c 242 g 340 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.47e-156 Length: 1197
 Score: 1680.00 Matches: 310
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-10-070-464-3 (1-310) x AF221635 (1-1197)

QY 41 HisCysAspPhePheIleSerIleYrSerAsnGlnIysAsnProHisCysValSerLeu 60
 DB 123 CACTGTGACTCTCTTTAATAAGTAAGTATAGTACAGAAAGATCCAGCTGTGTCCTT 182
 QY 61 TyrIleuSerSerProGluAspAspProThrCysIysThrIysGluPheTPrAlaThr 80
 DB 183 TACAACTATCAAGTCCCTGAAGATGACCCAACTTGCAAAACAAAGAAATTTTGGGCACC 242
 QY 81 IleuAspSerIleagIyProleuProAspTyrThrProGluIlePheSerPheGlu 100
 DB 243 ATTTGGATTACAGAGGCTCTCTTCGACTATACCTCCAGAAATTTCTTTGAA 302
 QY 101 SerThrThrIysPheThrLeuTyrGlyMetLeuTyrIysProHisAspLeuGlnProGly 120
 DB 303 AGTACTACTGATTTTACATTGTATGGATGCTCTCAGAACGCTCATGATCAGACCTCGA 362
 QY 124 LysIyPTrProThrValIleuPheIleTyrGlyGlyProGlnGlnIleGluIleAsp 140
 DB 363 AAGAAATATCTACTGCTGCTGCTCATATATGATGCTCAGAGGTCAAATAGAAATTGAC 422
 QY 141 AspGlnValIglIyLeuGlnIyTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160
 DB 423 GATCAGGTGGAAGGACTCCAAATATCTAGCTTCTGATATATATTCATTGACTGATCGT 482
 QY 161 ValGlyIleHisIagIyTSPserTyrGlyIyTyrIleuSerIleuMetAlaLeuMetGlnArg 180
 DB 483 GTGGGCAATCCAGCGCTGTCTTATGGAGATACCTCTCCGTAGGCAATTATGACAGG 542
 QY 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTPrIlePheTyrAsp 200
 DB 543 TCAGATATCTTCAGGGTGTGCTATGCTGCGGGCCCACTGCTGTGATCTTCATGAT 602
 QY 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnIyTyrTyrIleu 220
 DB 603 ACAGGATACAGGAGCTTATATGGTACCTGACAGATGACAGGGCTTATTACTTA 662
 QY 221 GlySerValAlaMetGlnIagIyLysPheProSerGluProAsnArgLeuLeuLeu 240
 DB 663 GGATCTGTGGCAATGACAGAGAAAGTCCCTGACCAACCAATCGTTATCTGCTCTTA 722
 QY 241 HisGlyPheIleuAspGluAsnValHisPheAlaHisThrSerIleuLeuSerPheLeu 260
 DB 723 CATGGTTTCTCGATGAGATGATGTCATTTGCATATCCAGTATATTAATGAGTTTCTTA 782
 QY 261 ValArgIleagIyLysProTyrAspLeuGlnIleTyrProGlnGlnIyTyrHisSerIleArg 280
 DB 783 GTAGGAGCTGGAAGCCATATGATTTTCAAGTCTATCTCAGAGAGCACAGCATTAAGA 842
 QY 281 ValProGluSerGlyGlnHisTyrGlyLeuHisIleuLeuHisTyrLeuGlnIyAsnLeu 300
 DB 843 GTTCCGTAATCGGAGAACTTATGAACTGATCTTTGGACATCTTCAAGAAACCTT 902
 QY 301 GlySerArgIleAlaIleuLeuValIle 310
 DB 903 GGATCAGCTATGTGCTCTTAAGTGATA 932
 RESULT 2
 AX354795 2649 bp DNA linear PAT 06-FEB-2002
 LOCUS AX354795
 DEFINITION Sequence 3 from Patent WO0179473.
 ACCESSION AX354795
 VERSION AX354795.1 GI:18619528
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 Meyers,R.A. and Williamson,M.
 21953, a human prolyl oligopeptidase family member and uses thereof
 Patent: WO 0179473-A 3 25-OCF-2001:
 Millennium Pharmaceuticals, Inc. (US)

```

FEATURES
    source
        1..2649
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT      803 a      514 c      586 g      746 t
ORIGIN
Alignment Scores:
Pred. No.:      9.63e-153      Length:      2649
Score:          1645.50      Matches:      310
Percent Similarity: 86.35%      Conservative: 0
Best Local Similarity: 86.35%      Mismatches: 0
Query Match:    97.95%      Indels:      49
DB:             6      Gaps:      1
US-10-070-464-3 (1-310) x AX354795 (1-2649)
QY      1  PheGluGlyThrIlyAspSerProLeuGluHisIleuTyValIleSerTyValAsn 20
DB      1570 TTGAAAGGACCAAGACTCCCTTTAGAGCATCCTGTAGTACGATGATTAAGTAAAT 1629
QY      21  ProGlyValIleThrArgLeuThrAspArgGlyTySerHisSerCysIleSerGln 40
DB      1630 CCGAGAGGAGGACAGGCTGACTGACCGGTGCTACCTCATTCTTGCATGATCAGTCAG 1689
QY      41  HisCysAspPhePheIleSerTySerAsnGlnIlyAsnProHisCysValSerIleu 60
DB      1690 CACTGTACTCTTTATTAAGTAATAGTATGATACAGAAATCCACTGTGTCTCCCTT 1749
QY      61  TyrIlySerSerProGluAspAspProThrCysIlyThrIlyGluPheThrIleThr 80
DB      1750 TACAAAGCATCAAGTCCGAAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACCC 1809
QY      81  IleuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
DB      1810 ATTTGGATTCAGAGAGGCTCTTCCTGCACTACTCTCCAGAAATTTCTCTTTGAA 1869
QY      101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyIlyAspProHisAspLeuGlnProGly 120
DB      1870 AGTACTACTGGATTACTTGTATGGAGTCTTCAACAGCTCTCATGATCTACAGCTTGA 1929
QY      121 LysIlyTyProThrValIleuPheIleTyGlyIleProGln----- 134
DB      1930 AAGAAATATCTACTGCTGTCTCATATATGTGTGCTCTCAGTGACGATTGTGTAAT 1989
QY      134 ----- 134
DB      1990 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTAGCCTCTAGATTATGTGTT 2049
QY      134 ----- 134
DB      2050 GTACTGATAGACAAAGGGATCTGTTCACGAGGGCTTAATTTGAAGGGCCTTTAAA 2109
QY      135 -----GlyGlnIleGluIleAspAspGlnValGluIlyLeuGlnTyIleuAlaSer 151
DB      2110 TATTAATATGGTCAATATAGAAATGACGATCAGGTGGAAGGACTCCAAATATCTAGCTTCT 2169
QY      152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTTPSerTyrGlyIlyTyr 171
DB      2170 CGATATGATTTCACTTGAATGATGCTGTGGGCTCCACGCGCTGCTCTTATGAGGATAC 2229
QY      172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB      2230 CTCTCCCTGATGCGATTATGACAGAGGTCAATATCTTCAGGGTGTCTATTTCTGGGACC 2289
QY      192 ProValThrLeuTyrIlePheTyAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
DB      2290 CCAAGTCACTCTGTGATCTTCTATGATACAGATACCGGAACCTTATATGGGTCACTCCT 2349
QY      212 AspGlnAsnGlnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluIlyPhePro 231
DB      2350 GACCAGAAATGAACGGGCTATTAAGTCTGTGGCCATGCAAGAGAAAGTTTCCCTC 2409

```

```

QY      232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB      2410 TCTAACCAATCTTTACTGCTCTTACATAGTGTTCCTCGATGAGATGTCATTGTGCA 2469
QY      252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIlyProTyrAspLeuGlnIle 271
DB      2470 CATACAGTATATTAAGTATTTTATGATGAGGCTGGAAAGCCATATGATTACAGATC 2529
QY      272 TyrProGlnIlyuArgHisSerIleArgValProGluSerGlyIleHisIleGluHis 291
DB      2530 TATCTCAGAGAGACACAGCATTAAGTCTCTTAATCGGAGGAACATTATGAATGATCAT 2589
QY      292 LeuLeuHisTyIlyLeuGlnGluAsnLeuGlySerArgGlyIleAlaIleLeuValIle 310
DB      2590 CTTTGACCTACTCTTCAAGAAACCTTGATCAGATGATGCTGCTTAAAGTGATA 2646

RESULT 3
AY172659
LOCUS      2656 bp      mRNA      linear      PRI 08-JAN-2003
DEFINITION Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPP1)
ACCESSION AY172659
VERSION    AY172659.1
KEYWORDS   mRNA, complete cds.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2656)
            Makaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            Q1.S., Akinsanya,K., Riviere,P. and Junien,J.-L.
            Novel Serine Protease Genes Related To DPP1V
            Patent: US (WO 0231134)-A 18-Apr-2002;
            2 (bases 1 to 2656)
            Q1.S., Akinsanya,K., Riviere,P. and Junien,J.-L.
            Direct Submission
            Submitted (04-NOV-2002) Perring Research Institute, 3550 General
            Atomics Ct., San Diego, CA 92121, USA

FEATURES
    source
        1..2656
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /chromosome="15"
        /map="15q22.1-q22.2"
        1..2656
        /gene="DPP1"
        8..2656
        /gene="DPP1"
        /codon_start=1
        /product="dipeptidyl peptidase IV-related protein-1"
        /protein_id="AA017261.1"
        /db_xref="GI:27549550"
        /translation="MAAMETEOAGVEIFETADCEENIESQDRPKLEPFYVERYSMSQ
        LKLLADTRKYGMAKAPHDPMFVFNPDGPHSRIRYLLAMSGNRENTLISEI
        PKTINRAVAVLMSMKPLDLFOATLPGWMSREBELREKRTCTGCTAAYDYQSGG
        TIFPOAGSGVYHKXDGPOFPTQQLRPNUVETSCNPTRNDPKLCPADPMWATFHSN
        DIVISNIVREERLLTYVNELANMEDDASAGATFVLQEEFDRIYSGYVWCPAAETT
        PASGKILRLIYENDSESEVEI IHVTSPEMLTBRADSRFRYKGTANRKYFKSEIM1
        DAGRGILIDVDKELIOPFELIFEGVEYIARAGWPBGKYVMSILLDSOTRLQVLIS
        PELFIPVEDDVMRORLIEVPSVPLIIEETTDIWINIHIFHFVPGSHSEIEIF
        IFASECKTGPRHLYKITSLSKSYRSGGLPAPSPFKPIKEAATGSEMEVLCR
        HGSNIQVDERRLVYREGTDSLEHLYVVSIVNPEVRLTDRGTSKSCISQHC
        FPLSKYSNQNPICVSLYKISSPEDPTCKTKEFWATILDSAGPLPYTPPELFSFS
        TTFPTLYGMILPKHDLPQPKYPTVLFIVGQVQVLYVNNPFGVKYFRMLNTLISLGV
        VVIVIDNRGSHRGLKEGAFKRYNGQIEIDDVGGQYLAARYDFIDLDBGYHGMGY
        GGYLSMALMORSDFRVAIAGAPVTLMIYDVGTYERVYMGHPDQNGGYLISVAMQ
        AEKPESEPNRLLIHGRIDENVRHARTSILSLFVAAKRYDILQITVQENHSIRVPS
        GEHYEHLHLHYLOENTGSRALAKVT"
BASE COUNT      804 a      517 c      588 g      747 t
ORIGIN
Alignment Scores:

```

Pred. No.: 9.66e-153 Length: 2656
 Score: 1645.50 Matches: 310
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 0
 Query Match: 97.95% Indels: 49
 DB: 9 Gaps: 1

US-10-070-464-3 (1-310) x AY172659 (1-2656)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyValSerTyValAsn 20
 DB 1577 TTGAAGGACCAAGACTCCCTTTAAGCATACCTGTAAGTACGTAATTAAT 1636
 QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTySerHisSerCysAlleserGln 40
 DB 1637 CCTGGAGAGTGACAAAGCTGACTGACCGGTGCTACTCACAATTCCTTGTCATCATGCTGAG 1696
 QY 41 HisCysAspPhePheIleSerLysTySerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1697 CACTGTGACTCTTTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 1756
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 80
 DB 1757 TACAAGCTATCAAGTCTGAAGATGACCAAGCTGCAAAACAAAGAAATTTGGGCCACC 1816
 QY 81 IleuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100
 DB 1817 ATTTTGATTCAGACAGGTCTCTCTCTGATATATCTCTCCAGAAATTTCTCTTTTGA 1876
 QY 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyLysProHisAspLeuGlnProGly 120
 DB 1877 AGTACTACTGATTTACATTTGTATGGAGTGTCTCAAGGCTCATGATCTACAGCTTGA 1936
 QY 121 LysLysTyProThrValLeuPheIleTyGlyGlyProGln----- 134
 DB 1937 AAGAAATATCTCTACTGCTCTCTCATATATGTGTCTCTCAGGTGCACTGGTAATAT 1996
 QY 134 ----- 134
 DB 1997 CGATTTAAAGAGTCAAGTATTCGCTTGAATACCTAGCCTCTAGTATATGTGTT 2056
 QY 134 ----- 134
 DB 2057 GTAGTATAGACAAAGGGGATCCTGTACAGGAGGCTTAATTGAAGGGCCTTTAA 2116
 QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLysLeuAlaSer 151
 DB 2117 TATAAAATGGTCAAAATGAATGACATGACGTGGAAGACCTCCAAATATCTAGCTTCT 2176
 QY 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTPSerTyGlyGlyTy 171
 DB 2177 CGATATGATTTCAATGACTTATGATCGTGTGGGCACTCCAGGCTGTCTATGAGAGTATC 2236
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 DB 2237 CTCCTCCGATGAGCATTTATATGACAGAGTCAAGATATCTTCAGAGGTGCTATGTCGGGGCC 2296
 QY 192 ProValThrLeuThrIlePheTyAspThrGlyTyThrGluArgTyMetGlyHisPro 211
 DB 2297 CCACTACTCTGTGATCTTCTATGATATCAGATATCAGAACTGTATATGTGGTCACTCCT 2356
 QY 212 AspGlnAsnGlnGlnGlyTyTyLysGlySerValAlaMetGlnAlaGluLysPhePro 231
 DB 2357 GACACAGATGAACAGGGCTATTTACTTATGATCTGTGGCCATGCAAGCAAGAAAGTTCCCC 2416
 QY 232 SerGluProAsnAspLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
 DB 2417 TCTTAACCAAAATCGTTTACTCTCTTACATGTGTTTCTCGATGAGATGTCATTTTGA 2476
 QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnIle 271
 DB 2477 CATACAGATATATATGAGTTTATAGTGTGGGCTGGAAGGCAATATGATTTACAGATC 2536

QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyGluLeuHis 291
 DB 2537 TATCTTCAGAGACACAGCATTAAGTCTCTGAATCGGAGAACTATATGAATGATGAT 2596
 QY 292 LeuLeuHisTyLysGlnGlnAsnLeuGlySerArgIleAlaIleLeuValIle 310
 DB 2597 CTTTTCACCTACTCTTCAAGAAACCTTGATCATGATATGCTGCTTAAAGTGATA 2653

RESULT 4

AX608725 LOCUS 2671 bp DNA linear PAT 17-FEB-2003
 DEFINITION Sequence 2 from Patent WO0231134.
 VERSION AX608725
 KEYWORDS AX608725.1 GI:28404298
 SOURCE
 ORGANISM Homo sapiens (human)

REFERENCE
 AUTHORS Q1.S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.
 TITLE Novel serine protease genes related to dpp1v
 JOURNAL Patent: WO 0231134-A 2 18-Apr-2002;
 Ferring BV (NL)

FEATURES
 source location/Qualifiers
 1..2671
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 805 a 524 c 594 g 748 t
 ORIGIN

Alignment Scores:

Pred. No.: 9.72e-153 Length: 2671
 Score: 1645.50 Matches: 310
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 0
 Query Match: 97.95% Indels: 49
 DB: 6 Gaps: 1

US-10-070-464-3 (1-310) x AX608725 (1-2671)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyValSerTyValAsn 20
 DB 1577 TTGAAGGACCAAGACTCCCTTTAAGCATACCTGTAAGTACGTAATTAAT 1636
 QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTySerHisSerCysAlleserGln 40
 DB 1637 CCTGGAGAGTGACAAAGCTGACTGACCGGTGCTACTCACAATTCCTTGTCATCATGCTGAG 1696
 QY 41 HisCysAspPhePheIleSerLysTySerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1697 CACTGTGACTCTTTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 1756
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 80
 DB 1757 TACAAGCTATCAAGTCTGGAAGATGACCAACTGCAAAACAAAGAAATTTTGGGCCACC 1816
 QY 81 IleuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100
 DB 1817 ATTTTGATTCAGACAGGTCTCTCTCTGACTATATCTCTCCAGAAATTTTCTTTTGA 1876
 QY 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyLysProHisAspLeuGlnProGly 120
 DB 1877 AGTACTACTGATTTTACATTTGTATGGAGTGTCTCAAGAGCTCATGATCTACAGCTTGA 1936
 QY 121 LysLysTyProThrValLeuPheIleTyGlyGlyProGln----- 134
 DB 1937 AAGAAATATCTCTGCTGTTCATATATGTGTGCTCAGGTGCACTGGTAATAT 1996
 QY 134 ----- 134
 DB 1997 CGATTTAAAGAGTCAAGTATTCGCTTGAATACCTAGCCTCTAGGTTATGTGTT 2056

QY 134 ----- 134
DB 2057 GTAAGTATAGACAACAGGGATCTGTCAACGAGGCTTAATTGAAAGCGCTTTAA 2116
QY 135 ----- G1yGln1leglu1leAspAspGlnVal1glu1yleuGln1rleuAla1Ser 151
DB 2117 TATAAATGGGTCAAAATAGAAATTTGACATCAGTGGAAAGGATCCCAATATCTAGCTTCT 2176
QY 152 ArgTyrAspPhe1leAspLeuAspArgVal1Gly1leH1sg1YTrpSerTyrGlyGlyTyr 171
DB 2177 CGATATGATTTTCATTAAGCTTAGATCGTGGCATCCACGGCTGGTCTTATGAGAGATAC 2236
QY 172 LeuSerLeuMetAla1leuMetGlnArgSerAsp1lePheArgVal1Ala1leAlaGlyAla 191
DB 2237 CTCTCCCTATAGGCACTTAATAGCAGAGGTCAAGATATCTTCAGGGTTGCTATTCGTGGGCC 2296
QY 192 ProValThrLeuTrp1lePheTyrAspThrGly1YTrThrGluArgTyrMetGlyH1sPro 211
DB 2297 CCAGTCACTCTGGATCTTCTTATGATACAGATACCGAAGCCTTATATGGGTCAACCT 2356
QY 212 AspGlnAspGlnGlu1ng1YrTyrLeuGlySerVal1AlaMetGln1AlaGlyLysPhePro 231
DB 2357 GACCAAGATGAAACAGGCTATTAAGTATGAGTCTGTGGCCATGCAAGCAAGAAAGTTCCCC 2416
QY 232 SerGluProAsnArgLeuLeuLeuLeuH1sg1YpHeuAspGluAsnVal1H1sPheAla 251
DB 2417 TCGAACCAAAATGTTTACTGCTCTTAACATGCTTCTCGATGAGAAATGTCATTTTGA 2476
QY 252 H1sThrSer1leLeuLeuSerPheLeuVal1Arg1Gly1LysProTyrAspLeuGln1le 271
DB 2477 CATACAGATATATTAATGATTTTATGAGGGCTGGAAGCCATATGATTTACAGATC 2536
QY 272 TyrProGlnGluArgH1sSer1leArgVal1ProGluSerGly1GluH1sTyrGluLeuH1s 291
DB 2537 TATCTCAGAGAGACACAGCATTAAGAGTCTTGAAATCGGGAACCATTAATGAACGTCAT 2596
QY 292 LeuLeuH1sTyrLeuGlnGluAsnLeuGlySerArg1leAla1leLeuLysVal1le 310
DB 2597 CTTTTCACACTTACTTCAAGAAACCTTGATCAACGATATTCCTCTTAATAAGTGAAT 2653
RESULT 5
AX342633 3106 bp DNA linear PAT 12-JAN-2002
LOCUS AX342633 Sequence 30 from Patent WO0198468.
DEFINITION AX342633
ACCESSION AX342633
VERSION AX342633.1 GI:18152030
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Yue,H., Elliott,V.S., Gandhi,A.R., Lal,P., Au-Young,J.,
Tribouley,C.M., Deleage,A.M., Baughn,M.R., Nguyen,D.B., Lee,E.A.,
Hafellia,A., Khan,F.A., Walla,N.K., Yao,M.G., Lu,D.A., Paterson,C.,
Tang,Y.T., Walsh,R.T., Azimzai,Y., Rankumar,J., Xu,Y. and Reddy,R.
JOURNAL Patent: WO 0198468-A 30 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
source 1..3106
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7160544CB1"
BASE COUNT 928 a 633 c 704 g 841 t
ORIGIN
Alignment Scores:
Pred. No.: 1,16e-152 Length: 3106
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0

Query Match: 97.95% Indels: 49
DB: 6 Gaps: 1
US-10-070-464-3 (1-310) x AX342633 (1-3106)
QY 1 PheGluGlyThrLysAspSerProLeuGluH1sH1sLeuTyrVal1SerTyrVal1Asn 20
DB 1772 TTTGAAAGGACCAAGAGCTCCCTTAGAGCACCACCTGACATGACATGACATGACAT 1831
QY 21 ProGluGluVal1ThrArgLeuThrAspArgGly1YrSerH1sSerCysGly1SerGln 40
DB 1832 CCTGGAGAGGTGACAAAGGCTGAGTCAACCGGTCTACATTCCTGCTGATCAGTCAAG 1891
QY 41 H1sCysAspPhePhe1leSerLysTyrSerAspGlnLysAsnProH1sCysVal1SerLeu 60
DB 1892 CACTGAGACTCTTTATTAAGTAAAGTATAGTAACAGAGATACACATGATGTCCTT 1951
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80
DB 1952 TACAGCTATCAAGTCTGAGATGAGTCAACCACTTGCAAAACAAAGGATTTGGGCCACC 2011
QY 81 H1sLeuAspSerAlaGlyProLeuProAspTyrThrProGlu1lePheSerPheGlu 100
DB 2012 ATTTTGATTCAGCAGAGTCTCTTCTGACTATACCTCCAGAAATTTTCTCTTTGAA 2071
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProH1sAspLeuGlnProGly 120
DB 2072 AGTACTAGTGAATTTACATGATGAGATGCTTACAGGCTCATGATGATACAGCTGGA 2131
QY 121 LysLysTyrProThrVal1leuPhe1leTyrGly1YrProGln----- 134
DB 2132 AAGAAATATCCATCTGCTGCTTCAATATGATGGTCTCAGGTGACATGTCGAATAT 2191
QY 134 ----- 134
DB 2192 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTCGT 2251
QY 134 ----- 134
DB 2252 GTAAGTATAGACAACAGGGATCTGTCAACGAGGCTTAATTGAAAGCGCTTTAA 2311
QY 135 ----- G1yGln1leglu1leAspAspGlnVal1glu1yleuGln1rleuAla1Ser 151
DB 2312 TATAAATGGGTCAAAATAGAAATTTGACATCAGTGGAAAGGATCCCAATATCTAGCTTCT 2371
QY 152 ArgTyrAspPhe1leAspLeuAspArgVal1Gly1leH1sg1YTrpSerTyrGlyGlyTyr 171
DB 2372 CGATATGATTTTCATTAAGCTTAGATCGTGGCATCCACGGCTGGTCTTATGAGAGATAC 2431
QY 172 LeuSerLeuMetAla1leuMetGlnArgSerAsp1lePheArgVal1Ala1leAlaGlyAla 191
DB 2432 CTCTCCCTATAGGCACTTAATAGCAGAGGTCAAGATATCTTCAGGGTTGCTATTCGTGGGCC 2491
QY 192 ProValThrLeuTrp1lePheTyrAspThrGly1YrThrGluArgTyrMetGlyH1sPro 211
DB 2492 CCAGTCACTCTGGATCTTCTATGATACAGATACACGGAACGTTATATGGTCAACCT 2551
QY 212 AspGlnAspGlnGlu1ng1YrTyrLeuGlySerVal1AlaMetGln1AlaGlyLysPhePro 231
DB 2552 GACCAAGATGAAACAGGCTATTAAGTATGAGATCTGTGGCCATCAAGCAAGAAAGTTCCCC 2611
QY 2297 SerGluProAsnArgLeuLeuLeuLeuH1sg1YpHeuAspGluAsnVal1H1sPheAla 251
DB 2612 TCGAACCAAAATGTTTACTGCTTACATGGTTCCTCGAATGAGAAATTCATTTTGA 2671
QY 252 H1sThrSer1leLeuLeuSerPheLeuVal1Arg1Gly1LysProTyrAspLeuGln1le 271
DB 2672 CATACAGATATATTAATGATTTTATGAGGGCTGGAAGCCATATATTTACAGATC 2731
QY 272 TyrProGlnGluArgH1sSer1leArgVal1ProGluSerGly1GluH1sTyrGluLeuH1s 291
DB 2732 TATCTCAGAGAGACACAGCATTAAGATTTCTGATCGGAGAACATTAATGAATGCAAT 2791

QY 292 LeuLeuHisTyrLeuGlnGluValenLeuGlySerArgIleAlaIleValIle 310
 Db 2792 CTTTGACACTACCTTCAAGAAACCTTGATCAGCATGCTGCTCTAAAGTGATA 2848

RESULT 6
 AF221634
 LOCUS AF221634 3127 bp mRNA linear PRI 05-NOV-2000
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
 ACCESSION AF221634
 VERSION AF221634.1 GI:11095187
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 3127)
 REFERENCE 1 Abbott,C.A., Yu,D.M., Woollett,E., Sutherland,G.R., McCaughan,G.W.
 and Gorell,M.D.
 Cloning, expression and chromosomal localization of a novel human
 dipeptidyl peptidase (DPP) IV homolog, DPP8
 Eur. J. Biochem. 267 (20), 6140-6150 (2000)
 MEDLINE 20467194
 PUBMED 11012666
 2 (bases 1 to 3127)
 REFERENCE 2 Abbott,C.A., Yu,D.M., McCaughan,G.W. and Gorell,M.D.
 Direct Submission
 Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,
 Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia
 JOURNAL Location/Qualifiers
 FEATURES
 source
 1..3127
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q22"
 1..3127
 /gene="DPP8"
 214..2862
 /gene="DPP8"
 /note="serine protease; member of prolyl oligopeptidase S9b
 family belonging to enzyme clan SC; alternatively spliced
 forms of this mRNA are found in AF221635, AF221636 and
 AF221637"
 /codon_start=1
 /product="dipeptidyl peptidase 8"
 /protein_id="AAC29766.1"
 /db_xref="GI:11095188"
 /translation="MAAAMEEQLGVEIFETADCEENIESQDRKLEPFYVERYSMSQ
 LKULADTRKHGIMAKAPADHFWVRNDPGHSDRIYGLMNGENRENTLFSSEI
 PKTINRAVALMLSKPLDLFQATLDYGMYSREBELRERKRIQVGIADYHQSG
 TPLFOAGSGIYHVDGPGFTQOPLRNLYVETSCPIRMDPKICPADPMIAFIHSH
 DIWISNIVTSEERLTYVHNELAMMEDASAGATVLDSEEPGSGYGMCPKAEET
 PSGKILRILEENDESEVEIHYTSPMLSTRADSRFRYKGTGNAKPKPFKSEMI
 DAGSRITDIVDKELIOPEYLFEVEYIAAGTPEBKAWKSILDRSOTRLQIVLIS
 PELFIPEVDVMEKRLIESVPSVTPILIIYEETDWINIHIFHPFQSHSEELER
 IFASCEKTRGRHLKYITSLIKESKYKSSGLPAPSDPKPIKEBAITGEBEVLGR
 HGSNIQVDEVRALVYFEGTDSPLRHLVYVSVNPGEVRLIDRGYSHSCCSQHD
 FFLSKYNQKPHCVSLYKLSBEDPTCKTEFMATILDSAGLPLPYTPPEIFSES
 TTGETLYGLMYKPHDLOGKKYPTVLFITYGPOVLVNNRPFKGVKRYRLNTLSLGV
 VVIVINDGSGHGLKPGKAGKVKMGQIEIDDOVGLQYASRYDILDDIVGHSY
 GYLISLMALMORSDFRVALAGAPVTIMITYDGYTERTYRIGHFPONSQGYLDSVAHQ
 AEKFPSEPNRLLLHGFLDENVHFAHTSILLSPVRAGKPYDLQIYPERHSIRVPS
 GEHYELHLHYLDENLGSRIALKVI"

BASE COUNT 943 a 637 c 706 g 841 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,17e-152 Length: 3127
 Score: 1645.50 Matches: 310
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 0
 Query Match: 97.95% Indels: 49

DB: 0 9 Gaps: 1
 US-10-070-464-3 (1-310) x AF221634 (1-3127)

QY 1 PheGluGlyThrTyrAspSerProLeuGluHisIleLeuTyrValIleSerTyrValAsn 20
 Db 1783 TTTGAAGGACACCAAGACTCCCTTTAGACATACACTGATGACGATTCGTAAT 1842

QY 21 ProGlyValuValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
 Db 1843 CTGGAGAGGTGACCAAGCTGACGACCGGCTACTACATCTTCTGTCATCAGCAG 1902

QY 41 HisCysAspPheIleSerIleTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
 Db 1903 CACTGTGACTCTTTATTAAGTATAGTATAGTACAGAAATCCACACTGTGTCCTT 1962

QY 61 TyrIleLeuSerSerProGluAspAspProThrCysIleThrIleGluPheTyrPalaThr 80
 Db 1963 TACAGCTATCAAGTCTTGAAAGTGAACCAACTTGCAAAAGAAATTTGGGCTACC 2022

QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
 Db 2023 ATTTGGATTCAAGAGCTCTCTTCTGACTATACCTCTCAAGAAATTTCTCTTGAA 2082

QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleAspLeuGlnProGly 120
 Db 2083 AGTACTACTGATTTACATTGATGATGATGATGATGATGATGATGATGATGATGAT 2142

QY 121 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGln 134
 Db 2143 AAGAAATATCTACTGCTGCTGCTCATATATATGATGATGATGATGATGATGAT 2202

QY 134 134
 Db 2203 CGGTTTAAAGAGTCAAGTATTTCCGCTGAATACCTACCTCTAGGTTATGCTGTT 2262

QY 134 134
 Db 2263 GTAGTATAGACAACAGGAGCTCTGTCACCGAGGCTTAATTTGAAGGCGCTTTAAA 2322

QY 135 135
 Db 2323 TATTAATGGGCTCAATATGAATTTAGCATGATGATGATGATGATGATGATGATGAT 2382

QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyTyr 171
 Db 2383 CGATATGATTTCACTTGAATCTAGATCTGATGATGATGATGATGATGATGATGAT 2442

QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 Db 2443 CTCTCCCTGATGGCATTAAAGCAGAGGTGATATATCTTCAGGGTTGCTATGCGGGCC 2502

QY 192 ProValThrLeuProIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 Db 2503 CCACTACCTCTGATCTTCTATGATACAGATACAGATACAGATACAGATACAGATAC 2562

QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
 Db 2563 GACCAAGATGAACAGGCTATTAATCTAGATCTGTGGCCATGACAGCAAGAAAGTTCCC 2622

QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluValHisPheAla 251
 Db 2623 TCTGAACCAAAATGTTTACTGCTCTTACATGATGATGATGATGATGATGATGATGAT 2682

QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
 Db 2683 CATACAGATATATTAATGATGATTTTATGATGATGATGATGATGATGATGATGATGAT 2742

QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerIleGlyHisTyrGluLeuHis 291
 Db 2743 TATCCTCAGAGAGAGACACAGCATTAAGATTCCTGATTCGGAGAAACATTATGAATGAT 2802

QY 292 LeuLeuHisTyrLeuGlnGluValenLeuGlySerArgIleAlaIleValIle 310

Db 2803 CTTTGACACTACCTTCAAGAAAACCTTGATCAGTATGCTCTTAAGATGATA 2859

RESULT 7
AX354793 3143 bp DNA linear PAT 06-FEB-2002

LOCUS AX354793
DEFINITION Sequence 1 from Patent WO0119473.
ACCESSION AX354793
VERSION AX354793.1 GI:18619526

KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
1 Meyers, R.A. and Williamson, M.
21993, a human prollyl oligopeptidase family member and uses thereof
Patent: WO 01/9473-A 1 25-OCT-2001;
Millennium Pharmaceuticals, Inc. (US)

FEATURES
source
1. 3143
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
229. 2877
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD22990.1"
/db_xref="GI:18619527"

CDS
/translation="MAAMETSQLGVEIFETADCEINIESODRPLEPFYVERYSMO
LKLADLRKHYHMAKAPHDPMFVKRNDPDGSDHLYLAMGENRENTLFSYSEI
PTIRAAVILMLMKPLDLFOATLDYGMYSREBLERKIGVGLASVYHGGSG
TFIFQAGSIVYKDGQPGFOPLRPLVETSCPNIRMDPKLPADDMATFHSN
DIMSNIVTERRRLTYVNEILAMNEBARSAGVATPULOEPDSYCYMCKPKEET
PSGKILILYBENDESVEITIHVTSPLMETRADSPRYPKGTIANPPTKMSGITI
DAEGRIIDVIDKELIQFEILFEGVEYLARAGMTPEGKYMILIDRSQTRQIVLIS
PELFIPEDDVNERQLIESVDSVTPLIYEETDIDINHDIFHVPQSHSEIER
IFASCKTGFRHLKYITSLIKESKYRSQGLPAPSDPCPKKEIAITSGSEVTLR
HGSNIQVDEVRLLVYEGTKDPLBHLIYVVSYPAGEVTRLTDRGSHSCCISOCH
FPIKSYNOKNPKHCVSLYKLSPEDDPKCTKEPMATLIDAGPLPDYTPPEIPES
TTGFTLYKMLYKRPDLOPKRPTLYLFIYGGQYQVLYNNRFGKVFPLANTLASGIV
VVVIDNRSCNRGLKFEAGFKYKQDIEIDVDEGLQYLASVYDFTIDIDRGVHGMSTY
GGYLSLMLMORSDIFRAVIAAGPVLMIYDGTGTERYMGHDPDNEGQYLGVSAMQ
AAKPESENRLLHLGFLDENYHFAHTSLILSLFVRAKPKYDQLIYQPERHSIRVPS
GEHYELHLIYQENLNGSRIMALKVI"

BASE COUNT 943 a 644 c 712 g 844 t

ORIGIN

Alignment Scores:
Pred. No.: 1.18e-152 Length: 3143
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
Gaps: 1

US-10-070-464-3 (1-310) x AX354793 (1-3143)

QY 1 PheGluGlyThrIlyAspSerProLeuGluIuHsiIseuTyYvaIvalSerTyYtalasn 20
Db 1798 TTGGAAGACACCAAGACTCCCTTTAAGCATCAGCTAGTGAAGTAACTAAAT 1857

QY 21 ProGluGluValThArGLeuThzArpArgLyTySerHisSeCySeIleSerGln 40
Db 1858 CTTGAGAGGTGACAAAGCTGACTGACCTGCTACTCATCTTCTGCTGACTGCTCAG 1917

QY 41 HisCyAspPhePheIleSerIlyTySerAsnGlnIlyAsnProHisCyAsIleSerLeu 60
Db 1918 CACTGTGACTCTTATATAGTAAGTAAACCAAGAAATCCACACTGTGTGCTCTT 1977

QY 61 TyrlsIleuSerSerProGluAspAspProThrCysIlyThrIlyGluIuPheIalThr 80
Db 1978 TACAAGTATCAAGTCTGGAAGATGACCAACTTGCAAAAGAAATTTTGGCCACC 2037

QY 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProProGluIuIlePheSerPheGlu 100
Db 2038 ATTTGGATTTCAGCAGGCTCTTCTTCTGACTATCTCTCCCAAGAAATTTCTTTGAA 2097

QY 101 SerThrThrGlyPheThrIleuTyGlyMetIleuTyLysPheProHisAspLeuGlnProGly 120
Db 2098 AGTACTACTGATTTACATTTGATGAGATGCTCTCAAGCCCTCAATGATCTACAGCTGGA 2157

QY 121 LysIlyTyProThrValLeuPheIleTyGlyGlyProGln----- 134
Db 2158 AAGAAATATCCATCTGTCTGTTCATATATGATGATGCTCCAGATGACATGATATAT 2217

QY 134 ----- 134

Db 2218 CGTTTAAAGAGTCAAGATTTCCGCTTGAATACCCTAGCTCTAGCTATGTGCTT 2277

QY 134 ----- 134

Db 2278 GTATGATAGACAAAGGAGATCTGTCAAGAGGCTTAAATTTGAAGCGGCTTTAA 2337

QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSer 151
Db 2338 TATTAATATGAGTCAATATGAAATTCAGATCAGTGGAGGATCCCAATATCTAGCTTCT 2397

QY 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyGlyGlyTyYr 171
Db 2398 CGATATGATTCATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2457

QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
Db 2458 CTCTCCGATGAGGATTAATGAGAGGTCAGATATCTTCAAGGCTTGCTATCTGGGCGC 2517

QY 192 ProValThrLeuThrIlePheTyAspThrGlyTyThrGluArgTyMetGlyHisPro 211
Db 2518 CCAGTCACTCTGTGATCTTCTATATATACAGATACAGGATACAGGATATATATGATGAT 2577

QY 212 AspGlnAsnGluGlnGlyTyTyLeuGlySerValAlaMetGlnAlaGluIlyAspPhePro 231
Db 2578 GACCAATATGACAGGCTATTTACTTACAGATCTGAGGATGACAGGATGACAGGATGAT 2637

QY 232 SerGluProAsnArgLeuLeuLeuLeuIleGlyPheLeuAspGluIlyAsnValHisPheAla 251
Db 2638 TCTGAACCAATTCGTTTACTGCTCTTACATGATGTTCTCGATGATGATGATGATGAT 2697

QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIlyTyTyAspLeuGlnIle 271
Db 2698 CATACAGTATATATATGATTTTATGATGAGGCTGAGAAAGCATATGATATACAGATC 2757

QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluIlyHisIleHis 291
Db 2758 TATCTCAGAGAGACACAGATTAAGATTCCTGAAATGGGAGAAATTAATGAACTGCAT 2817

QY 292 LeuLeuHisTyLeuGlnGluIlyAsnLeuGlySerArgIleAlaIleLeuIlyValIle 310
Db 2818 CTTTGACACTACCTTCAAGAAAACCTTGATCAGTATGCTCTTAAAGATGATA 2874

RESULT 8
BC040203 4535 bp mRNA linear PRI 03-DEC-2002
LOCUS BC040203
DEFINITION Homo sapiens, similar to dipeptidyl peptidase 8, clone MGC:42570
IMAGE:4824813, mRNA, complete cds.
ACCESSION BC040203
VERSION BC040203.1 GI:26007916
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 4535)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2002) National Institutes of Health, Mammalian

REMARK
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabba-remail.nih.gov
Tissue Procurement: Miklos Palakovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHRG1) & Shiroki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnarsson, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H., Kowars, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IPAK Plate: 71 Row: k Column: 23.

FEATURES
SOURCE

1. 4535
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:42570 IMAGE:482413"
/cissue_type="Testis"
/clone_id="NTH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
/codon_start=1
/product="similar to dipeptidylpeptidase 8"
/protein_id="AAH40203.1"
/db_xref="GI:2607917"
/db_xref="LOCUSID:54878"
/translation="MAAAMETEQLEVEIFETADCEINIESODRPLEPFYVERYSWSQ
LKLLADTRKYHYMMAKAPHDPMFKNDPDGPHSDRIYVLAMSGENRENTLPSYSEI
PKTINRAVLMLSMKPLDLFOATLDYVMSREBELERKIGVGIASDYQSG
TFLPQAGSIVHVKDQSGPOGFTQOPLRNVLVETSCPNRMDPKCPADPMIAFVHGN
DIWISNITRREBRLLTYVHNLAMNEEDASAGVATFLQSEPRYSYMMCPRAET
PSGKILAILYEENDESEVEITHYSPMLERRADSPRYRTGTANPRVTYRKSEINI
DAGRIIDVIRELQPELIPBGEVITARGMTPBEGTASLIDRSQTLQIVLIS
PELFPVEDDVMERQLIESVPSVTPLIYEETDIIINIHDIHFHPQSEETIER
IFASECKTFPHLYKITSILKESKYKSSGGLPAPSDKCPIKEBIAITSGEVELGR
HGSINQVDEVRLLVYFEGTKDSPLEHLLVYVYVNGEVTLLDDBGYSHSCISQCHD
FEISKYSNOKNPHCVSLYKLSSPEDDPCTCKEPMATILDSAGPLPDYTPREISFES
TTGRTLYGMLYKPHDLQPGKPYTLPIYGQPOVQVUNNRKQKYPRLNTLASLGYV
VVVIDNRGSCRGLKFEBAFYKMGQIIDDQVEGLQYLAIRYDITIDRVLGHSWT
GGYLSLMLMORSDIFRVALGAPVTLMI FYDTGYTERYMGHPDNEGYLGVAMQ
AEKPESENRLLHLHGFLDENVHPMTSILSLFLVRACKPYDLQIYPOERSHIVPES
GEHYELHLHYLOENIGSRIALAKVI"

CDS

BASE COUNT 1313 a 918 c 946 g 1358 t
ORIGIN
Alignment Scores:
Pred. No.: 1.81e-152 Length: 4535
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 49
Query Match: 97.95% Indels: 49
DB: 9 Gaps: 1
US-10-070-464-3 (1-310) x BC040203 (1-4535)

Qy 1 PheGluGluThrIleAspSerProLeuGluHisIleLeuTyValValIserTyrValasn 20
Db 3198 TTTGAAGCACCAAGACTCCCTTTAGGATCCCTGTACGTAGTACGTACCTAAT 3257
Qy 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
|||||

Db 3258 CTTGAGAGGTGACAGAGGCTGACCGCTGGCTACTACATTTCTTCTGCTGATCAGTCAg 3317
Qy 41 HisCysAspPhePheIleSerLeuTyrSerAsnGlnIleAspProHisCysValSerLeu 60
Db 3318 CACTGTGACTTCTTTATAGATAGTATAGTACAGAGAAATCCACATGTGTGCTCTT 3377
Qy 61 TyrIleAspSerSerProGluAspAspProThrCysValThrIleGluPheThrAlaThr 80
Db 3378 TACAAGCATCAAGTCCGAGAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACC 3437
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
Db 3438 ATTGTGATTCAGCAGGCTCTTCTGCTGACTATCTCTCCAGAAATTTCTCTTTGAA 3497
Qy 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleProHisAspLeuGlnProGly 120
Db 3498 AGTACTACTGATTTACTATTATAGGATGTCTTCAAGAGCTCATGATCTACAGCTTGA 3557
Qy 121 LysIleTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
Db 3558 AAGAAATATCTACTGCTGCTGCTCATATATGTGTGTCTCCAGTGCAGTGTGATTAAT 3617
Qy 134 ----- 134
Db 3618 CGGTTTAAAGATCAAGTATTTCCGCTGAATACCTAGCCTCTTATGTATGTGTT 3677
Qy 134 ----- 134
Db 3678 GTAGTATAGACAAAGGCGATCTCTGCACCGAGGCGCTAAATTTGAAGGCGCTTTAA 3727
Qy 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
Db 3728 TATTAATATGGTCAAAATGAAATGACATCAGTGCAGTGCAGTGCATATCTAGCTTCT 3797
Qy 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPserTyrGlyGlyTyr 171
Db 3798 CGATATGATTTTCATGACTTATGATGCTGTGGGATCCACGCGTGTCTTATGAGAGATAC 3857
Qy 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
Db 3858 CTCCTCCGATGAGGATTAATGACAGATGATATCTTCAGGTTGCGATTCGAGGCGCC 3917
Qy 192 ProValThrLeuThrPhePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
Db 3918 CCAGTCACTCTGTGATCTTCTTATGATACAGATACACGGAACGTTATATGGGTACACCT 3977
Qy 212 AspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLeuPhePro 231
Db 3978 GACCAAGATGAACAGGCTATTAATGATCTGTGCGCATGACAGCAAGAAAGTTCCCC 4037
Qy 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
Db 4038 TCTGAACCAAAATCGTTTACTCTCTTACATGTGTTCTCGATGAGAAATGCTCATTTGCA 4097
Qy 252 HisThrSerIleLeuAspSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
Db 4098 CATACCAATATTAATGATGATTTTATGATGAGGCTGAGAAACCATATGATTTACAGATC 4157
Qy 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGlyLeuHis 291
Db 4158 TATCTCAGAGAGACACACGATTAAGACTTCTGTAATCGGAGAAACATTATGAATGAT 4217
Qy 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuValIle 310
Db 4218 CTTTTCACATCACTTCAAGAAACCTTGATGATCAATGATGCTGCTCAAAAGTATA 4274
RESULT 9
LOCUS AX608743 4676 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 20 from Patent WO0231134.
ACCESSION AX608743
VERSION AX608743.1 GI:28404307
KEYWORDS

SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 Qi, S., Akinsanya, K.O., Riviere, P.J. and Junien, J.L.			
TITLE	Novel seven protease genes related to dplyl			
JOURNAL	Patent: WO 02/31134-A 20 18-Apr-2002;			
FEATURES	Ferring BV (NL)			
source	Location/Qualifiers			
	1..4676	/organism="Homo sapiens"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"		
BASE COUNT	1424 a	859 c	979 g	1414 t
ORIGIN				
Alignment Scores:				
Pred. No.:	2,35e-152	Length:	4676	
Score:	1644.50	Matches:	308	
Percent Similarity:	99.35%	Conservative:	0	
Best Local Similarity:	99.35%	Mismatches:	1	
Query Match:	97.89%	Indels:	2	
DB:	6	Gaps:	1	
US-10-070-464-3 (1-310) x AK608743 (1-4676)				
QY	1	PheGluGIYTrhLYAspSerProLeuGIuHSHiSLeuTYrValIaSerTYrValaSn	20	
Db	1783	TTTGAAGCACCMAAGACTCCCTTTAGACATCACCTGTACGTAGTCACTTAAT	1842	
QY	21	ProGIyGIuValThraXLeuThraSpaRgGIYrSerHiaserCysIleSerGI	40	
Db	1843	CCTGAGAGGAGCACAAGCCTGACTACCGGTGGCTACACTTCTTGTCGATCAGT	1902	
QY	41	HisCysAspPhePheIleSerIlyrSerAnGIuLYaSPProHiasCysValSerIu	60	
Db	1903	CACGTGTACTCTTTATAGTAGTAGTAGTAACCAAGAAATCCACACTGTGTCCCTT	1962	
QY	61	TYrLYsLeuSerSerProGIuAspAspProThrCysLYsTrhLYsGIuPheThraIaThr	80	
Db	1963	TRCAAGCTATCAAGCCCTGAAGATACCCCACTTCCAAAACMAAGATTTGGGCCACC	2022	
QY	81	IleLeuAspSerIaGIyProLeuProaSPTYrThrProProGIuIlePheSerPheGIu	100	
Db	2023	ATTTTGATTCAGCAGGTCTCTTCTCCTGACTATACTCTCCAGAAATTTCTCTTTGAA	2082	
QY	101	SerThrThrGIyPheThrLeuTYrGIyMeLeuTYrLYsProHiasPLeuGIuProGIy	120	
Db	2083	AGTACTACTGCAATTACATTGATGGATGCTTCAAGCCCTCATGATCTACAGCTGGA	2142	
QY	121	LYsLYsTYrProThrValLeuPheIleTYrGIyGIyProGIuNGIuIleGIuIleAsp	140	
Db	2143	AAGAAATATCTTACTACTGTCTGTCTTCAATATGGTGTGCG---GGTCAATAGAAATTTGAC	2198	
QY	141	AspGIuValGIyGIyLeuGIuTYrLeuHiaserArgTYrAspPheIleAspLeuAspArg	160	
Db	2199	GATCAGGTGGAAGGACTCCAATATCTTACTCTTCTGATATGATTTCACTTATGATGCT	2258	
QY	161	ValGIyIleHiasGIYTPSerTYrGIyGIyTYrLeuSerLeuMeCalaLeuMeGIuArg	180	
Db	2259	GTTGGGCAATCCACGGCTGTCTCTATGAGAAATACCTCTCCATGGCCATTATATCAGAGG	2318	
QY	181	SerAspIlePheArgValaIaIleIaIaGIyValaProValThrLeuTrpIlePheThyAsp	200	
Db	2319	TCAATATCTTCAAGGTGCTATTCCTGTGGGCCCAAGTCACTCTGTGGATCTTTATGAT	2378	
QY	201	ThrGIyTYrThrGIuArgTYrMeGIyHiasProaSPGIuAsnGIuNGIyTYrTYrLeu	220	
Db	2379	ACAGAGATACAGGAACGTTATATGGGTCAACCTCAACGATATGAACAGGGCTATTACTTA	2438	
QY	221	GIySerValaIaMeGIuIaGIuLYsPheProSerGIuProaSnArgLeuLeuLeu	240	

Dp	2439	GGATCTGTGGCCATGCAAGCAAGAAAAGTTCCTCCCTGACCAAAATCGTTTACTGCTCTTA	2498
Qy	2481	HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu	260
Dp	2499	CATGCTTTCCTCGATGTAGAAATGATGCTCATTTTGACATACCAATATATTACGAGTTT	2558
Qy	261	ValArgAlaGlyIleProTyrAspLeuGlnIleTyrProGlnGluAaGHisSerIleArg	280
Dp	2559	GTCAGGGCTGGAAAGCCATATGATTTACAGATCTATCTCTCAGAGACACACGATAGA	2618
Qy	281	ValProGluSerGlyGluHisTyrGluLeuHisLeuHisTyrLeuGlnGluAsnLeu	300
Dp	2619	GTTCCCTGATCGGAGAACATTAATGAACCTGCATCTTTTGCACTACCTTCAGAAACCTT	2678
Qy	301	GlySerArgIleAlaIleLeuValIle	310
Dp	2679	GGATCAGCTATTGCTGCTCTTAAAGTGATA	2708
RESULT 10			
AA608737			
LOCUS	AX608737	4309 bp	DNA
DEFINITION	Sequence 14 from Patent WO0231134.		linear
ACCESSION	AX608737		
VERSION	AX608737.1	GI:28404304	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	Qi, S., Atkinsanya, K.O., Riviere, P.J. and Junin, J.L.		
	Novel serine protease genes related to dppiv		
	Patent: WO 0231134-A 14 18-APR-2002;		
	Ferring BV (NL)		
FEATURES			
source	Location/Qualifiers		
	1..4309		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	1304 a 790 c 907 g 1308 t		
ORIGIN	0		
Alignment Scores:			
Pred. No.:	1.62e-150	Length:	4309
Score:	1625.50	Matches:	310
Percent Similarity:	85.87%	Conservative:	0
Best Local Similarity:	85.87%	Mismatches:	0
Query Match:	96.76%	Indels:	51
DB:	6	Gaps:	1
US-10-070-464-3 (1-310) x AX608737 (1-4309)			
Qy	1	PheGluGlyThrIleAspSerProLeuGluHisLeuTyrValValSerTyrValAsn	20
Dp	1263	TTTGAAAGGCACCAAGACTCCCTTTAGACATCACTGACGATACGATACGTAAT	1322
Qy	21	ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln	40
Dp	1323	CCTGAGAGGGGAGACAGCTGACGACCGGCGGCTACCAATCTTGCTGCATCAAGTCAG	1382
Qy	41	HisCysAspPhePheIleSerTyrTyrSerAsnGlnIleAsnProHisCysValSerLeu	60
Dp	1383	CACGTGACTCTTTAATAGTAAGTATAGTACCAAGAAATCCACACTGTGTCTCTT	1442
Qy	61	TyrIleuLeuSerSerProGluAspAspProThrCysIleThrIleGluPheTrpAlaThr	80
Dp	1443	TACAAAGCTATCAAGTCTCGTAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCACC	1502
Qy	81	IleLeuAspSerAlaGlyProLeuAspProTyrThrProGlnIlePheSerPheGlu	100
Dp	1503	ATTTTGATTCACGAGGTCTCTTCCGACTATACCTCCCAAAATTTTCTCTTTGAA	1562
Qy	101	SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleSerProHisAspLeuGlnProGly	120

```

Db      1563 AGTACTACGATTTTACATTGATGAGATGCTCTACAGGCTCATGATCTACAGGCTTGA 1622
Qy      121 LVSLVETYPProThrValLeuPheIleTYRGLYGLY--ProGln----- 134
Db      1623 AAGAAATATCTACTGTGCTTGCATATATGAGGGTCTCTCAGGTGAGCTGTGTAATA 1682
Qy      134 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYRLeuAlaS 151
Db      1683 ATCGGTTTAAAGAGTCAGATATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGG 1742
Qy      134 ----- 134
Db      1743 TTGTAGTATAGACAACAGGGATCTGTGCACCGAGGCTTAATTGAAAGGCGCTTTA 1802
Qy      135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYRLeuAlaS 151
Db      1803 AATATATAATGGGTCAATAATGAAATGACGATCAGGTGGAAGGACTCCAAATATCTAGCTT 1862
Qy      151 eFArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTPSerTYRGLYGLY 171
Db      1863 CTGGAATATGATTTCACTTGAATGATCGTGTGGCATCCAGGCTGGTCTTATGAGAGAT 1922
Qy      171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 191
Db      1923 ACCTCTCCCTGATGAGCATTTATGTCAGAGGTCAAGATATCTTCAGGGTTGCTATTGCTGG 1982
Qy      191 lAProValThrLeuTPRlePheTYrAspPheTYrThrArgIleuArgTYrMetGlyHisP 211
Db      1983 CCCCATCTCTCTGTGTGATCTTCTATGATACAGGATACAGGAACGTTATATGGGTACCC 2042
Qy      211 rAspGlnAsnGlnGlnGlyTYRTYRLeuGlySerValAlaMetGlnAlaGluLysPheP 231
Db      2043 CTGACCAAGATGAACAGGGCTATTACTTAGAGATCTGTGGCATGCAAGCAAGAAAGTTC 2102
Qy      231 rOsSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251
Db      2103 CCTCTAACCAAAATCGTTTACTCTCTTACATGATGTTTCCGTGATGAGAAATGTCATTGTTG 2162
Qy      251 lAHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAspLeuGlnI 271
Db      2163 CACATACCGATATATATCTAGATTTTATGAGGGCTGGAAGCCATATGATTTACAGA 2222
Qy      271 lEYrProGlnIleuArgHisSerIleArgValProGlnSerGlyGluHisTYRGLYLeuH 291
Db      2223 TCTATCTCTGAGGAGAGCAACAGCATTAAGATTCCTGAATCGGAGAAACATTATGAATCGC 2282
Qy      291 lELeuLeuHisTYRLeuGlnIleuAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310
Db      2283 ATCTTTTGACCTACCTTCAGAAAAACCTTGATCAGGATTCGTGCTTAAAGTGATA 2341

```

```

RESULT 11
AX608735 4829 bp DNA linear PART 17-FEB-2003
LOCUS AX608735
DEFINITION Sequence 12 from Patent WO0231134.
ACCESSION AX608735
VERSION AX608735.1 GI:28404303
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Qi,S., Akinaanya,K.O., Riviere,P.J. and Junien,J.L.
AUTHORS Novel serine protease genes related to dppiv
JOURNAL Patent: WO 0231134-A 12 18-APR-2002;
Featuring BV (NL)
FEATURES
Location/Qualifiers
source 1..4829
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1466 a 886 c 1017 g 1460 t

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1,85e-150 Length: 4829
Score: 1625.50 Matches: 310
Percent Similarity: 85.87% Conservative: 0
Best Local Similarity: 85.87% Mismatches: 0
Query Match: 96.76% Indels: 51
DB: 6 Gaps: 1

```

US-10-070-464-3 (1-310) x AX608735 (1-4829)

```

Qy      1 PheGlnGlyThrLysAspSerProLeuGlnHisLysLeuTYrValIleSerTYrValAsn 20
Db      1783 TTTGAAAGCACCAGAAAGACTCCCTTAGAGCATCACCTGATGACTAGTACGTAAT 1842
Qy      24 ProGlyGluValIleThrArgLeuThrAspArgGlyTYRSerHisSerGlySerIleSerGln 40
Db      1843 CCGGAGAGGTGCAAGGCTGACTGACCGTGGCTACTCAACATTTGTCGATCAGTCAG 1902
Qy      41 HisCysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisCysValSerLeu 60
Db      1903 CACTGTGACTCTCTTTATAGTAATATGATTAACAGAGATTCACACTGTGTCTCTT 1962
Qy      61 TYrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 80
Db      1963 TACAAGCTATCAAGTCTGTAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCACC 2022
Qy      81 lLeuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePheSerPheGlu 100
Db      2023 ATTTGGATTACAGCAGGCTCTCTTCCTGACTATACCTCCAGAAATTTTCTTTTGA 2082
Qy      101 SerThrThrGlyPheThrLeuTYrGlyMetCLeuTYrLysProHisAspLeuGlnProGly 120
Db      2083 AGTACTAGGATTTTCAATTGATGAGGATGCTTACAGACCTATGATCTACAGCCTGGA 2142
Qy      121 LVSLVETYPProThrValLeuPheIleTYRGLYGLY--ProGln----- 134
Db      2143 AAGAAATATCTCTACTGTGCTTCAATATATGGTGGTCTCTCAGGTGAGGAGTGTGTAATA 2202
Qy      134 ----- 134
Db      2203 ATCGGTTTAAAGAGTCAGATATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGG 2262
Qy      134 ----- 134
Db      2263 TTGTAGTATAGACAACAGGGATCTGTGCACCGAGGCTTAATTGAAAGGCGCTTTA 2322
Qy      135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTYRLeuAlaS 151
Db      2323 AATATATAATGGGTCAATAATGAAATTGACGATCAGGTGGAAGGACTCCAAATATCTAGCTT 2382
Qy      151 eFArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTPSerTYRGLYGLY 171
Db      2383 CTGGAATATGATTTCACTTGAATGATCGTGTGGCATCCAGGCTGGTCTTATGAGAGAT 2442
Qy      171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 191
Db      2443 ACCTCTCCCTGATGAGCATTTATGACAGATCAGATATCTTCAGGCTGGTCTATGCTGGGG 2502
Qy      191 lAProValThrLeuTPRlePheTYrAspPheTYrThrArgIleuArgTYrMetGlyHisP 211
Db      2503 CCCCACTCACTCTGTGATCTTCTATGATACAGATACAGGAACGTTATATAGGTGACCC 2562
Qy      211 rAspGlnAsnGlnGlnGlyTYRTYRLeuGlySerValAlaMetGlnAlaGluLysPheP 231
Db      2563 CTGACCAAGATGAACAGGGCTATTACTTAGAGATCTGTGGCCAGCAGCAAGAAAGTTC 2622
Qy      231 rOsSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251
Db      2623 CCTCTAACCAAAATCGTTTACTGCTCTTACATGATGTTTCCGTGATGAGAAATGTCATTGTTG 2682
Qy      251 lAHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAspLeuGlnI 271

```


D6	2693	CACATACAGATATATTACTGAGTTTATTAGTAGGAGCGTGAAGGCATATGATTACACGA																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
----	------	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```

source
1. .4799
   /organism="Mus musculus"
   /mol_type="mRNA"
   /strain="C57BL/6"
   /db_xref="taxon:10090"
   /clone="MGC:58057 IMAGE:6410075"
   /tissue type="Brain, enriched mouse brain 12.5dp"
   /clone_id="NIH_BMAP_F00"
   /lab host="DH10B"
   /note="Vector: pTX-ASC"
1. .4799
   /gene="Dpp8"
   /gene="Locustid:74388"
   /db_xref="Locustid:74388"
   /db_xref="MGI:1921638"
288..2966
   /codon_start=1
   /product="Dpp8 protein"
   /protein_id="AAH43124.1"
   /db_xref="GI:27695450"
   /db_xref="Locustid:74388"
   /translation="MKIPSGRCMAAAMETEOLGVEIFETAECEGNGESGDPRKLEP
FYVERYSQOLTKLADTRKYHGYMAKAQHDPMFVPRPDGPHSRVUYVLMSGEN
RENTLRFSELPKTIINRAVALMSWKPLDLPQALTDGYMRSREELREKRIGTVGI
AAVDYHPSGGLFQAGSGIYHIDKGDPGKPTQPPNIVETSCPIRADPKLCPND
PMDIAFLHSDIWIISNLVTRSEERRITYVNHETLANMEEDPRSGAVATVLOEEDRYSYG
YMCQPAERPSGSKILRIYIENDESEVEIIVTSPMETRRADSRFYKGTGANPK
VTPEMSIIVDDAAGIIVDIDKELOVEFELFEGVEYIAAGWTECKHMSIILDSF
OHTLQVLSIPELFPIVEDAMRORLIIVESVSYPLIYEETDQIMNIIHIFNPF
POTHEDEIERIPASECTGCFRHHYKITSILKEKRYKSSGGLPARSPFKPIKEETIT
TSGEVAGRHGSNIWDEARKLVYFEGIKDSFLENHLVYTSTANPEEVRLTDRGYS
HSCCLSHDCDFISKYSNQNPRCVSLYKLSPEDDPVHRTKEFMATLDSAGLPDY
TPEPIEFSESTGTFTLYGMLYKPHDLPGKKYPTVLFIYCGQVOLVNNPKGVKPYR
LNTLASIGYVVVYIDNRGSCHRGLEKFGAKRYMGQIEIDDOVGLOYLASQVDFDL
DRYIGHGMSYGYLSIIMALMORSDIFPVALIAGA PVLMIIFYDTGYTERWGHDPNQO
GYLGYSAMQAKKEKPEPSPPNRLILHGFIDENHGFAPHSIILSLVYAKGRYDQLIYQ
ERISIRPESGEYEHLLHYLDENSGRIALAKVI"

BASE COUNT      1373 a      1015 c      1103 g      1308 t
ORIGIN

Alignment Scores:
Pred. No.:      1,74e-148      Length:      4799
Score:          1605.50      Matches:      302
Percent Similarity: 84.96%      Conservative: 3
Best Local Similarity: 84.12%      Mismatches: 5
Query Match:    95.57%      Indels:      49
DB:             10      Gaps:      1

US-10-070-464-3 (1-310) x BC043124 (1-4799)

QY      1 PheGluGlyThrIlysAspSerProLeuGluHisIleuTyValValSerTyValAsn 20
Db      1887 TTTGAAAGGCACCAAGAAGCTCTCTTGAGACATCACTGACGACCAAGTATGCAAC 1946
QY      21 ProGluValValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIleSerGln 40
Db      1947 CCTGGAGAAGGTGAGAGCTGACGACCGGAGCTACCACTCCGCTGCTCAGGCG 2006
QY      41 HisCysAspPhePheIleSerIyTySerAsnGlnIysAsnProHisCysValSerLeu 60
Db      2007 CATGTGACTTCTTCATTAAGTAAAGTACAGCAACAGAAAGATCACACTGTGTGCTC 2066
QY      61 TyrIlyIleuSerSerProGluAspAspProThrCysAllyThrIlyGluPheTrrPALthr 80
Db      2067 TACAAACTCTCAAGTCTTGAGAGTACCCAGTTCATTAACAAAGAAATTTGGGCCACC 2126
QY      81 IleLeuAspSerAlaGlyProLeuProAspTyTrThrProProGluIlePheSerPheGlu 100
Db      2137 ATTTGGATTACAGAGTCTCTTCCGACTACACCCCTCAGAAATTTTCTTTTGA 2186
QY      101 SerThrThrGlyPheThrIleuTyrglyMetLeuTyIysProHisAspLeuGlnProGly 120
Db      2187 AGAAGTACTGAGATTTCACGTGTGAGAAAGTGTGTATTAAGCTCATGACCTACCAACCTGGA 2246

```



```

QY      121 LysLysTyrProThrValLeuPheIleTyrGlyProGln----- 134
DB      2247 AGAAATACCCCACTGTTATTCATATATGTCGTCGCCAGGTGACGTCGTGAACAAT 2306
QY      134 ----- 134
DB      2307 CGGTTAAAGAGTCAAGTATTTCCGCTGGAACACCTGGCTCCCTGGTTATGTGTT 2366
QY      134 ----- 134
DB      2367 GTGGTATAGACAACAGGGATCTCTGCACCGAGACTTAATTGAAAGCGCTTTAA 2426
QY      135 -----GlyGlnIleGlnIleAspAspGlnValGlnIleGlnIleGlnIleAsp 151
DB      2427 TATAAATGGGTCAATATAGAAATCGATGATCAAGTGAAGAGCTCCAGTACCTAGCATCT 2486
QY      152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyGlyTyr 171
DB      2487 CAGTATGACTTCAATGATCTTGATCGAGTGGGATCCACCGCTGCTTATGGTGGCTAC 2546
QY      172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB      2547 CTCTCCCTATAGGCACTTAATGACAGAGTCCGATATCTCCGGGTGGCTATGCTGGGGCC 2606
QY      192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlnArgTyrMetGlyHisPro 211
DB      2607 CCAGTACACCTCTGATCTTCTTATGATACAGATACACGAGCGCTATATGGCTCACCCCT 2666
QY      212 AspGlnAspGlnGlnIleGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
DB      2667 GACCAAGAAAGAACAGGCTCACTAATGATCTGTGGCCATGCAAGGCGGAAGTCCCC 2726
QY      232 SerGlnProAspArgLeuLeuLeuLeuHisGlyPheLeuAspGlnValHisPheAla 251
DB      2727 TCAGAACCAACCGGTTACTCTCTTGCATGGGTTCTGATGAGAAATGTTCACTTTGCA 2786
QY      252 HisThrSerIleLeuLeuSerPheLeuValAlaAlaGlyLysProTyrAspLeuGlnIle 271
DB      2787 CACACCAAGATATATGCTGATCTTTTATGATGAGGCTGGAAAGCCATATGCTTACAGATC 2846
QY      272 TyrProGlnGlnArgHisSerIleArgValProGlnSerGlyHisTyrGlnLeuHis 291
DB      2847 TATCTCAGAGAGGACAGACATCAAGTTCCTGAGTCTGGAACACCTATGAACTGGCAC 2906
QY      292 LeuLeuHisTyrLeuGlnGlnIleAsnLeuGlySerArgIleAlaLeuLysValIle 310
DB      2907 CTGCTCACACTTCACTTCAAGAGAACCTTGATCGCGTATCGCTCTGAAAGATGAT 2963
RESULT 13
LOCUS   BC030688 3130 bp mRNA linear PRI 30-MAY-2002
DEFINITION Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:26191
ACCESSION BC030688
VERSION   BC030688.1 GI:21265132
KEYWORDS MGC.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3130)
AUTHORS Strausberg, R.
TITLES Direct Submision
JOURNAL Submitted (24-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgabs-r@mail.nih.gov
          Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
          cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
          Toshiyuki and Piero Carninci (RIKEN)

```

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jeessia Fahey, Erin Helton, Mark Kerteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
0
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAX Plate: 33 Row: d Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18450277.
Location/Qualifiers
1. 3130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26191 IMAGE:4822550"
/issue_type="Test18"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
372..2867
/codon_start=1
/product="Similar to dipeptidylpeptidase 8"
/protein_id="AAH30688.1"
/db_xref="GI:21265133"
/translation="MAAAMEEOLGVEIFETADCEENIESODRPKLEPFVERYSMSO
LKULADTRKNGYMMAKAPHPMPYKNDPPGPHSDRYIILAMSENNENTLPSYSEI
PTKINPAAYLMSMKPLDLFOATLDYGNYSBEELRKRKIGTGLASVDHQSQ
TFLFOAGSGIYHVKKGGPQGFQOAPLPLVETSCINIMDPLCPADPDMAAFHISN
DIWISNIVTREERRLTYVNEELANNEEDRSGAVAFVLOEEDRYSGYWCKAETT
PGGKILRLIYENDESVEIILHTVSPLMETRADSRYPKGTGAPKPYKPLTASLSTI
DAEGRIIDVIDKEILOPELIFEGVYIARACMTPEGKXAMSIILDSORLOIVIS
PELRTIPVEDNVERORLTSVPSDSTPLIYSETDIYININDIFHYVPSHEEIEP
IRASBCKTFPHLKYITSLKSKTKRSGGLPABDFPCPIKEEATILSGEVEVGR
HOSNIQVDEVRRLVYEGTKDSPLEHRLVVSIVNGEVTRLDKGYSHSCCISQFD
FISKYSNKNPKHCVSLYLSSPEDDPCTKTEFMAITLDSAGPLDPYDPPEISFES
TTGFTLYGMLYRPHLDQPKKYPTVLFIYGGPOVLNNRPFQVKTFRNTLASLSTI
VVVIDRSGCHRGKBEKAFKRYVAIACAPVLMTFYDTGYTERYMGHPDQEOGY
LGSVMAOKPSEPRILLILGCFIDENYHFAHTSLISFLVRAGKPYLDQTPQGRH
STRVBSGCHYHLHLHYIQENUGSRIALAKVI"
BASE COUNT 955 a 626 c 717 g 832 t
ORIGIN
Alignment Scores:
Pred. No.: 4,49e-129 Length: 3130
Score: 1407.00 Matches: 270
Percent Similarity: 88.36% Conservative: 11
Best Local Similarity: 84.91% Mismatches: 19
Query Match: 83.75% Indels: 18
DB: 9 Gaps: 5
US-10-070-464-3 (1-310) x BC030688 (1-3130)
QY      1 PheGlnGlyThrLysAspSerProLeuGlnIleHisLeuTyrValIleSerTyrValAsn 20
DB      1941 TTGAAGGACCAACAAAGACTCCCTTAAAGCATCACTTATAGTATCGTAAAT 2000
QY      21 ProGlnGlnValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
DB      2001 CCGGAGAGGTACCAAGGCTGACCTGCTCACTCAATCTTGCGCATCACTGAC 2060
QY      41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAspProHisCysValSerLeu 60
DB      2061 CACTGCACTTCTTTATAGTAAAGTAAATCCAGAAAGTCCACACTGTGTCTCTT 2120
QY      61 TyrLysLeuSerSerProGlnAspAspProThrCysLysThrLysGlnPheTrpAlaThr 80
DB      2121 TCAAGGCTATCAAGTCTGAGATGATGCCAATTCGAAACAAAGAAATTTGGGCCACC 2180
QY      81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnIlePheSerPheGln 100

```

```

Db      2181 ATTTGGATTGACAGAGTCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 2240
Qy      101 SerThrThGlyPheThrLeuTyrglyMecLeuTyrlvsProHisAspLeuGlnProgly 120
Db      2241 AGTACTACTGAGATTTCATTGATGAGAGTCTCTACAGGCTCATATACAGCTTGGA 2300
Qy      121 LysLysTyPProThrValLeuPheIleTyrglyGlyPProGlnGlyGlnIleGluLeaP 140
Db      2301 AAGAAATATCTACTGCTGCTGCTCATATATGTTGGTCTCAGGTGCAAGTTG---GTGAAT 2357
Qy      141 AepGlnValGluGlyLeuGlnIlyr-----LeuAlaSer---ArgTyraP 154
Db      2358 AATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTACCTCTCTAGGTTATGTC 2417
Qy      155 PheIleAspLeuAspArgValGlyIle-----HisGlyTyrSerTyrglyTyrlsu 172
Db      2418 GTTGTATGATATACACACAGGGGATCTCTGACCCGAGGCTTAAATTTGAAGCGCTTT 2477
Qy      173 SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
Db      2478 AATATATAA-----ATGGTTGCTATTGCTGGGGCCCA 2510
Qy      193 ValThrLeuTrpIlePheTyraPThrGlyTyrlsuGlnArgTyrlsuMetGlyHisProAsp 212
Db      2511 GTCACTCTGTGATCTTCTATGATACAGGATACACGGAACGTTATATGGTCACTGAC 2570
Qy      213 GlnAsnGlnGlnGlyTyrlsuTyrlsuGlySerValAlaMetGlnAlaGlyLysPheProSer 232
Db      2571 CAGAAATGAACAGGGCTATTAATTAAGATCTGTGGCCATGCAAGCAAAAGTTCCCTCT 2630
Qy      233 GluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAlaHis 252
Db      2631 GAACCAAAATCGTTTACTCTCTTACATGGTTTCTGGATGAGATGTCATTGTCACAT 2690
Qy      253 ThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyraPLeuGlnIleTyrlsu 272
Db      2691 ACCAGTAAATTAATCTGATTTTAAAGAGGCTGGAAGCAATATGATTTACAGATCTAT 2750
Qy      273 ProGlnGlnArgHisSerIleArgValAlaProGlnSerGlyGlnHisTyrlsuGlnLeuHis 292
Db      2751 CCTCAGAGACACACAGCATTAAGAGTTCTGAAATCCGAGAACATTATGAACGCACTCT 2810
Qy      293 LeuHisTyrlsuGlnGlnAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310
Db      2811 TTGCACTACCTTCAAGAAACCTTGATCAGCATTTGCTCTTAAAGATGATA 2864

RESULT 14
BD157001 2161 bp DNA linear PAT 17-JAN-2003
LOCUS BD157001
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD157001.1 GI:27862759
VERSION JP 2002191363-A/11844.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2161)
Oca,T., Isogai,T., Nishikawa,T., Hayaashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K., and Otsuki,T.
Patent: JP 2002191363-A 11844 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/11844
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280900
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC

```

```

10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
PC Primer for synthesizing full-length cDNA and use thereof FH key
Location/Qualifiers
(16)..(1911).
FEATURES
source 1..2161
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 662 a 429 c 452 g 618 t
ORIGIN
Alignment Scores:
Pred. No.: 1,146-128 Length: 2161
Score: 1401.00 Matches: 269
Percent Similarity: 88.05% Conservative: 11
Best Local Similarity: 84.59% Mismatches: 20
Query Match: 83.39% Indels: 18
DB: 6 Gaps: 5
US-10-070-464-3 (1-310) x BD157001 (1-2161)
Qy 1 PheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyrlsuValSerTyraValAsn 20
Db 988 TTGAAAGGACCAAAAGACTCCCTTTAGGACATCACTGATGATGATGATGATGATGAT 1047
Qy 21 ProGlnGlnValThrArgLeuThrAspArgGlyTyrlsuHisSerCysCysIleSerGln 40
Db 1048 CCTGGAGAGGTACAAGGCTGACTGACCTGACTACTCATCTTCTGTCGATCACTGACG 1107
Qy 41 HisCysAspPhePheIleSerTyrlsuSerAsnGlnLysAsnProHisCysValSerLeu 60
Db 1108 CACTGATCTTCTTTATTAAGTAAAGTAAATCAACGAAAGATCCACATCTGTGCTCTT 1167
Qy 61 TyrlsuLeuSerSerProGlnAspAspProThrCysLysThrLysGlnPheTrpAlaThr 80
Db 1168 TCAAGGCTATCAAGCTCTGAATGATGACCACTTGCAAAACAAAGAAATTTGGGCGACC 1227
Qy 81 IleAspSerAlaGlyProLeuProAspTyrlsuProProGlnIlePheSerPheGln 100
Db 1228 ATTTGGATTGACAGGTCCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1287
Qy 101 SerThrThGlyPheThrLeuTyrglyMecLeuTyrlvsProHisAspLeuGlnProgly 120
Db 1288 AGTACTACTGAGATTTCATTGATGAGAGTCTCTACAGGCTCATATACAGCTTGGA 1347
Qy 121 LysLysTyPProThrValLeuPheIleTyrglyGlyPProGlnGlyGlnIleGluLeaP 140
Db 1348 AAGAAATATCTACTGCTGCTGCTCATATATGTTGGTCTCAGGTGCAAGTTG---GTGAAT 1404
Qy 141 AepGlnValGluGlyLeuGlnIlyr-----LeuAlaSer---ArgTyraP 154
Db 1405 AATCGTAAAGGAGTCAAGTATTTCCGCTTGAAATACCTTACCTCTCTAGGTTATGTC 1464
Qy 155 PheIleAspLeuAspArgValGlyIle-----HisGlyTyrSerTyrglyTyrlsu 172
Db 1465 GTTGTATGATATACACACAGGGATCTCTGACCCGAGGCTTAAATTTGAAGCGCTTT 1524
Qy 173 SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
Db 1525 AATATATAA-----ATGGTTGCTATTGCTGGGGCCCA 1557
Qy 193 ValThrLeuTrpIlePheTyraPThrGlyTyrlsuGlnArgTyrlsuMetGlyHisProAsp 212
Db 1558 GTCACTCTGTGATCTTCTATGATACAGGATACACGGAACGTTATATGGTCACTGAC 1617
Qy 213 GlnAsnGlnGlnGlyTyrlsuTyrlsuGlySerValAlaMetGlnAlaGlyLysPheProSer 232
Db 1618 CAGAAATGAACAGGGCTATTAATTAAGATCTGTGGCCATGCAAGCAAAAGTTCCCTCT 1677
Qy 233 GluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAlaHis 252

```

Db 1678 GAACAAATCGTTTACTGCTTACATGCTTCTCGATGAGATGTCATTTCACAT 1737
Qy 253 ThrSer11LeuLeuSerPheLeuValArgAlaGlyLeuProTyrAspLeuGln11eTyr 272
Db 1738 ACCAGTATATTACTGAGTTTCTTTAGTGAAGGCTGGAAGCCATATGATTACGATCTAT 1797
Qy 273 ProGlnGluValArgHisSer11eArgValProGlnSerGlyGlnHisTyrGlnLeuHisLeu 292
Db 1798 CCTCAGAGAGACACAGCATPAGAGATTCTCGAATCCGAGAACATTATGAACTGCATCTT 1857
Qy 293 LeuHisTyrLeuGlnGlnLeuValArgHisSer11eArgValAlaLeuValHis11e 310
Db 1858 TTGCACTACCTTCAGAAAACCTTGATCACTGATTCGCTCTTAAAGTGATA 1911
RESULT 15
AK027826 2161 bp mRNA linear PRI 01-ANU-2002
LOCUS Homo sapiens cDNA FLJ14920 fis, clone PLACE1007416, weakly similar
DEFINITION to D1PEPTIDYL PEPTIDASE IV (EC 3.4.14.5).
ACCESSION AK027826
VERSION AK027826.1 GI:14042789
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuna, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project
TITLE Unpublished
JOURNAL 2 (bases 1 to 2161)
REFERENCE Isogai, T. and Otsuki, T.
AUTHORS Direct Submission
JOURNAL Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction,
5' - & 3' -end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source Location/Qualifiers
1..2161
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1007416"
/cissue_type="placenta"
/clone_lib="PLACE1"
/note="Cloning vector: pME18SFL3"
16..1914
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55395.1"
/db_xref="GI:14042790"
/translation="MDPKLCPADPDWIAFIHSNDIWSINVTREERLLTYVNHENLNM
EDASAGVATFVLOEEDPRYSGYMPCPAETTPSGKILRIYENDSEVEIIVT
SMLETRRADSPRYKGTGTANPKVTGPMSEIMDAGR1DIVIDKELIQPELIPGV
EYIARAGMTPECKYWSIILDSORLQIVLISPELFIVEDDVWEROLISVDSV
TLLIYERTDTHINHDIFHPROSHIEIEIPASEKCTGRHLYKTSILSKSKY
KSSGGLPARSPDKPIKEI1AITSGEWVLGHSGNIVDVEVRLRVFEGTKSDPLE
HHLYVSYVNPGEVTRLDGRSYSHSCISQHDGDFLSKYGNKNPCVSLYKLSDE
DPCTCKEPMATILDSAGPLDYTPPEIIPSPSTGFTLYGMLYKPHDLQPKKPTV
LFTYGGPOVLNNRFGKGVKPYFLNTALAGVYVVVYIDRSGCHRLKEGAFKYMV
ATAGAPVTIMIFYDTGYTERWYGHGPNQNGYVLSVANQAKFSPERRLLLHGLFL

BASE COUNT 662 a 429 c 452 g 618 t
ORIGIN
Alignment Scores:
Pred. No.: 1,14e-128 Length: 2161
Score: 1401.00 Matches: 269
Percent Similarity: 88.05% Conservative: 11
Best Local Similarity: 84.59% Mismatches: 20
Query Match: 83.39% Indels: 18
DB: 9 Gaps: 5
US-10-070-464-3 (1-310) x AK027826 (1-2161)
Qy 1 PheGlnGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlaSerGln 20
Db 988 TTGAAAGGACCAAGACATCCCTTTAGAGCATCCTGTAGTACTGATGATGATTAAT 1047
Qy 21 ProGlnGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlaSerGln 40
Db 1048 CCTGAGAGAGTGAACAAGCTGATGACCTGATGATGATGATGATGATGATGATGAT 1107
Qy 41 HisCysAspPhePhe11eSerTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
Db 1108 CACTGACCTCTTTATTAAGTAAGTATGATTAACCAAGAAATCCACACTGTGTCTTT 1167
Qy 61 TyrLysLeuSerSerProGlnAspAspProThrCysValThrLeuGlnPheTyrAlaThr 80
Db 1168 TACAGCTATCAAGTCTTAAGATGACCCACTGCAACAAAGAAATTTGGGCCAC 1227
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGlnLysPheSerPheGln 100
Db 1228 ATTTGGATTCAGCAGGCTCTCTTCTGATGATGATGATGATGATGATGATGATGAT 1287
Qy 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
Db 1288 AGTACTACTGAGATTATCATTTGATGAGAGCTCTCAACCCATGATCTACAGCTGGA 1347
Qy 121 LysLysTyrProThrValLeuPheLeuTyrGlyGlyProGlnGlyGlnLysLeuAsp 140
Db 1348 AAGAAATATCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1404
Qy 141 AspGlnValGlnGlyLeuGlnTyr-----LeuAlaSer---ArgTyrAsp 154
Db 1405 AATCGGTTAAAGAGTCAAGATATTCGCTGGAATACCTTACCTGCTGATGATG 1464
Qy 155 Phe11eAspLeuAspArgValGlyLe-----HisGlyTyrSerTyrGlyTyrLeu 172
Db 1465 GTTGTAGTATGACAAACAGGGGATCTGTCAACGAGGCTTAAATTTGAAGCGCTTT 1524
Qy 173 SerLeuMetAlaLeuMetGlnArgSerAsp11ePheArgValAla11eAlaGlyAlaPro 192
Db 1525 AATATATAA-----ATGTTGCTATTCTGGGCCCA 1557
Qy 193 Val11eLeuTyr11ePheTyrAspThrGlyTyrThrGlnArgTyrMetGlyHisAspAsp 212
Db 1558 GTCACTCTGTGATCTTCTATGATACAGGATACAGGAACGTTATATGGGATCACCTGAC 1617
Qy 213 GlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlnLysPheProSer 232
Db 1618 CAGAAATGAACAGGGCATTAATCTAGGATCTGGCCATGCAAGCAAGAAAGTCCCTCT 1677
Qy 233 GluProAsnArgLeuLeuLeuLeuHis11eGlyPheLeuAspGlnLysValHisPheAlaHis 252
Db 1678 GAACAAATCGTTTACTGCTCTTACATGATTTCCGAGAGAAATGATTCATTTTGACAT 1737
Qy 253 ThrSer11eLeuLeuSerPheLeuValArgAlaGlyLeuProTyrAspLeuGln11eTyr 272
Db 1738 ACCAGTATATTACTGAGTTTCTTTAGTGAAGGCTGGAAGCCATATGATTACGATCTAT 1797
Qy 273 ProGlnGluValArgHisSer11eArgValProGlnSerGlyGlnHisTyrGlnLeuHisLeu 292

Db	1798	CCTCAGAGAGACACAGCATTAAGAGTTCCTGAATCGGAGACATTATGAACTGCATCTT	1857
Qy	293	LeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleuIysValIle	310
Db	1858	TTGCACCTACCTTCAGAAAACTTGGATCAGTATTGCTGCTCTAAAAAGTGATA	1911

Search completed: October 16, 2003, 00:04:20
Job time : 3407.49 secs

0

4

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 (Search time 227.774 Seconds

(Without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-3

Perfect score: 1680

Sequence: 1 FECTKDSPLEHLVYVYVN.....HLHLVQENLGRIRALKVI 310

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10070464/runac_15102003_113552_24799/app_query.fasta_1.2652
-DB=N_Geneseq_15Jun03 -QPM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pcc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464.@CGN 1.1.0 @runac_15102003_113552_24799 -NCPUs=6 -ICPU=3
-NO MAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THEREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCGAPXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_15Jun03:*

1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1680	100.0	1197	22	AAC85695	Nucleotide sequenc
2	1645.5	97.9	1821	24	ABV76411	Dipeptidyl peptida
3	1645.5	97.9	2643	24	ABH99935	Coding sequence of
4	1645.5	97.9	2649	25	ABX12255	CDNA encoding huma
5	1645.5	97.9	2671	24	ABK83322	CDNA encoding huma
6	1645.5	97.9	3106	24	ABK12892	Human protease PR
7	1645.5	97.9	3120	22	AAC85694	Nucleotide sequenc
8	1645.5	97.9	3120	24	ABD38956	Human dipeptidyl p
9	1645.5	97.9	3143	24	ABH99934	CDNA encoding 1195
10	1644.5	97.9	4676	24	ABK83328	CDNA encoding huma
11	1625.5	96.8	4309	24	ABK83328	CDNA encoding huma
12	1625.5	96.8	4829	24	ABK83327	CDNA encoding huma
13	1401	83.4	2161	22	AAH15009	Human CDNA sequenc
14	1377.5	82.0	1669	22	ABK85696	Nucleotide sequenc
15	1377.5	82.0	2668	24	ABN59775	Novel human coding
16	1377.5	82.0	2842	24	ABN59774	Novel human coding
17	1372	81.7	2510	24	ABD23843	Human protease PR
18	1358.5	80.9	4685	24	ABK83332	CDNA encoding huma
19	1352	80.5	4523	24	ABK83325	CDNA encoding huma
20	1094.5	65.1	3287	24	ABD38955	Alternative versio
21	1093.5	65.1	2415	24	ABD38957	Human dipeptidyl p
22	1093.5	65.1	2617	24	ABK83323	CDNA encoding huma
23	1093.5	65.1	2952	24	ABK69090	CDNA encoding huma
24	1093.5	65.1	3024	24	ABD38954	Human dipeptidyl p
25	1093.5	65.1	3716	24	ABG75955	Human PMK1 encodin
26	1093.5	65.1	4219	24	ABK83335	CDNA encoding huma
27	1093.5	65.1	4302	24	ABK83333	CDNA encoding huma
28	1089.5	64.9	3047	24	ABK83333	Nucleotide sequenc
29	1086.5	64.7	1083	22	AAC85697	CDNA encoding huma
30	1086.5	64.7	4180	24	ABK83339	CDNA encoding huma
31	1086.5	64.7	4263	24	ABK83338	CDNA encoding huma
32	1032.5	61.5	2027	21	AACT7137	Human ORFX ORF262
33	1032.5	61.5	2801	22	AA157896	Human polynucleoti
34	1032.5	61.5	2982	22	AA159666	Human polynucleoti
35	1032.5	61.5	3262	22	AA157880	Human polynucleoti
36	1024.5	61.0	2751	24	AAH38311	Murine dipeptidyl
37	934	55.6	973	22	AAH81719	Human protease and
38	924	55.0	925	24	ABE90148	Human polynucleoti
39	872.5	51.9	4076	24	ABK83337	CDNA encoding huma
40	872.5	51.9	4159	24	ABK83336	CDNA encoding huma
41	865.5	51.5	4037	24	ABK83341	CDNA encoding huma
42	865.5	51.5	4120	24	ABK83340	CDNA encoding huma
43	755.5	45.0	823	24	ABK30401	Human G-protein-co
44	708	42.1	3713	23	ABL10425	Drosophila melanog
45	708	42.1	3783	23	ABL06641	Drosophila melanog

ALIGNMENTS

RESULT 1
AAC85695
ID AAC85695 strand; cDNA; 1197 BP.
XX AAC85695;
AC AAC85695;
XX
DT 29-UTR-2001 (first entry)
DE Nucleotide sequence encoding human DPP8 524Phe-657Gln+707Gly-882Ile.
XX
XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
XX dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus; glucose intolerance;
KW immunosuppression; ss.
XX
XX Homo sapiens.
OS


```

PF 12-SEP-2000; 2000CN-0125127.
XX
XX 12-SEP-2000; 2000CN-0125127.
XX
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
XX WPI: 2002-529809/57.
XX P-PSDB; ABB99949.
XX
XX New dipeptidyl peptidase IV protein 27.5 and encoding polynucleotide,
XX useful for treating diabetes and tumors -
XX
XX Claim 6; Page 24-25 (Disclosure); 32pp; Chinese.
XX
XX The invention relates to dipeptidyl peptidase IV protein 27.5 (ABB99949)
XX and nucleic acids encoding it (ABV76411). The protein has a molecular
XX weight of 27.5 kD and has 40% identity and 58% homology over a 227 amino
XX acid stretch with Stenotrophomonas maltophilia dipeptidyl peptidase IV
XX (GenBank accession number D83263). The invention also relates to a method
XX for the recombinant production of the protein, an antagonist of the
XX protein, and the use of the protein, gene and antagonist in therapeutic
XX applications. Dipeptidyl peptidase IV protein 27.5 can be used in the
XX treatment of a variety of diseases such as diabetes and tumors. The
XX present sequence represents cDNA encoding dipeptidyl peptidase IV
XX protein 27.5.
XX
XX Sequence 1821 BP; 533 A; 374 C; 391 G; 523 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 4,766-184 Length: 1821
XX Score: 1645.50 Matches: 310
XX Percent Similarity: 86.35% Conservative: 0
XX Best Local Similarity: 86.35% Mismatches: 0
XX Query Match: 97.95% Indels: 49
XX DB: 24 Gaps: 1
XX
XX US-10-070-464-3 (1-310) x ABV76411 (1-1821)
XX
QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuTyValValSerTyValAsn 20
Db 499 TTTGAAAGCACCACCAAGACTCCCTTTAGAGCATCAGTGTAGTACGATTACCTTAAT 558
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
Db 559 CCGAGAGAGGTGACAAAGCTGACGCTGCTCTCACTCATTTCTTGGCTGACATCGTAG 618
QY 41 HisCysAspPhePheIleSerLysSerAsnGlnLysAsnProHisCysValSerLeu 60
Db 619 CACGTGTACTTCTTATAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 678
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 80
Db 679 TACAAGCATATCAAGCTCGAAGATGACCACTTCAAAAACAAAGAAATTTTGGGCCAC 738
QY 81 IleLeuAspSerAlaGlyProLeuProAspArgTyrThrProArgGluIlePheSerPheGlu 100
Db 739 ATTTGGATTCAGCAGGCTCTCTTCTGACTATACCTCCCAAGAAATTTTCTTTTAA 798
QY 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyLysProHisAspLeuGlnProGly 120
Db 799 AGTACTACTGATTTTACATGTATGAGATGCTTCACAAAGCTCAGATCTACAGCTGGA 858
QY 121 LysLysTyProThrValLeuPheIleTyGlyGlyProGln----- 134
Db 859 AAGAAATATCTACTGCTGTCTGTATATATGTTGTTCTCAGGTGACAGTTGTAATAT 918
QY 134 ----- 134
Db 919 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGTT 978
QY 134 ----- 134

```

```

Db 979 GTAGTATAGAACACAGGGAGTCTGTACCCAGGGCTTAAATTGAAAGCCCTTTAA 1038
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAsn 151
Db 1039 TATAAATGGGTCAATAGAAATTCACATCAGATCAGGTGAGAGCATCAATATCTACTTCT 1098
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyGlyGlyTyr 171
Db 1099 CGATATGATTTCTTACTGATTCATGATGTCGGGACATCCAGGCTGATCCTATGAGAGATAC 1158
QY 172 LeuSerLeuMetAlaLeuMetGlnAspSerAspIlePheArgValAlaIleAlaGlyAla 191
Db 1159 CTCTCCCTGATGGCATTTATGACAGAGGTCAAGATATCTCAGGGTTGCTATTCCTGGGCT 1218
QY 192 ProValThrLeuThrIlePheTyAspThrGlyTyrThrGluArgTyMetGlyHisPro 211
Db 1219 CCAGTCACTGCTGTGATCTTCTATGATACAGGATACAGGAACGTTATATGAGGTACCC 1278
QY 212 AspGluAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
Db 1279 GACCAAAATGAAACAGGGCTATTTACTTACGATCTGTGGCCATGCAAGAAAAAGTTCC 1338
QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
Db 1339 TCTGAACCAATTCGTTTACTGCTCTTACATGTTTCCGATGAGAAATGTCATTTTGA 1398
QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
Db 1399 CATACCACTATATTACTGATTTTAACTGAGAGGCTGGAAGCCATATGATTTACAGATC 1458
QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyGluLeuHis 291
Db 1459 TATCTTCAGAGAGACACAGCATTAAGCTTCTGATGAGGAGAACATTAATGAACGCAAT 1518
QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLysValIle 310
Db 1519 CTTTGGACTACCTTCAGAAACCTTGATCAGTATTTGCTGCTTAAAGTGATA 1575
XX
XX RESULT 3
XX AAH99935
XX ID AAH99935 standard; cDNA; 2643 BP.
XX
XX AC AAH99935;
XX
XX DT 12-APR-2002 (first entry)
XX
XX DE Coding sequence of 21953 human prolyl oligopeptidase.
XX
XX KW 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
XX cancer; cardiovascular disease; autoimmune disease; acopic allergy;
XX neuronal disorder; vascular disorder; prostate disorder; cytostatic;
XX antidiabetic; antiarthritic; antiasumatic; antiinflammatory;
XX diabetes mellitus; arthritis; multiple sclerosis; asthma;
XX Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
XX OS Homo sapiens.
XX
XX PN W0200179473-A2.
XX
XX PD 25-OCT-2001.
XX
XX PF 11-APR-2001; 2001WO-US40483.
XX
XX PR 18-APR-2000; 2000US-197508P.
XX
XX PA (MILL-) MILLENNIUM PHARM INC.
XX
XX PI Meyers RA, Williamson M;
XX
XX DR WPI: 2002-034353/04.
XX P-PSDB; AAG78415.
XX
XX

```


PT New polypeptides 21953, member of human prolyl oligopeptidase family,
 PT useful as diagnostic targets and therapeutic agents for controlling
 PT cancer, lymphoma and leukemia -
 XX
 PS Claim 2; Page-; 121pp; English.

XX This invention relates to an isolated 21953 human prolyl
 CC oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
 CC neuroprotective, antithyroid, dermatological, antipsoriatic,
 CC antineoplastic, ophthalmological, antiinflammatory, neurotropic,
 CC antiparasitomaian, anticonvulsant, gynecological, vasotropic,
 CC antitumoral, cardiant, antihypertensive, anorectic and
 CC metabolic in its action. Uses include gene therapy, expression or
 CC activity of 21953 protein modulator, it is useful for identifying a
 CC compound which binds to it and can be used in preventing, treating
 CC or detecting a cellular proliferative or differentiative disorder.
 CC The 21953 molecules can act as novel diagnostic targets and therapeutic
 CC agents for controlling disorders associated with the aberrant activity
 CC or degradation of peptide hormones e.g., disorders associated with cell
 CC differentiation and proliferation such as cancer, immune function,
 CC reproductive, neurological and cardiovascular function. The 21953
 CC molecules are thus useful for treating and preventing cellular
 CC proliferative and differentiative disorders, haematopoietic neoplastic
 CC disorders, immune disorders such as autoimmune diseases, diabetes
 CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
 CC neuronal disorders, demyelinating diseases, vascular disorders and
 CC metabolism or pain disorders. This sequence represents the cDNA
 CC encoding sequence of 21953 human prolyl oligopeptidase. This
 CC sequence represents the coding sequence of 21953, being the sequence
 CC in between the start and the stop codon of the sequence represented in
 CC AAH99934.

XX SQ Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;

Alignment Scores:

Pred. No.: 8, 37e-184 Length: 2643
 Score: 1645.50 Matches: 310
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 0
 Query Match: 97.95% Indels: 49
 DB: 24 Gaps: 1

US-10-070-464-3 (1-310) x AAH99935 (1-2643)

OY 1 PheGluGlyThrLeuAspSerProLeuGluHisLeuTyValValSerTyValAsn 20
 DB 1567 TTGAAGGACCAAGACCTCCCTTAGAGCATCACCTAGTACGTAAGTAAAT 1626
 OY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerGlySerGln 40
 DB 1627 CCTGGAGAGGTGCAAGGCTGACTGACCGTGGCTACTCACTTTGGCTGCACTGTCAG 1686
 OY 41 HisCysAspPhePheLeuSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1687 CACTGGATCTCTTTTAAGTAAGTATAGTAAACCAAGAACATCCACCTGCTGCTCTT 1746
 OY 61 TyrLysLeuSerSerProGluAspArgProThrCysValThrLysGluPheTyrPalaThr 80
 DB 1747 TACAAGCTATCAAGTCTCAAGATGACCAACTGCAAAACAAAGAAATTTGGGCCACC 1806
 OY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
 DB 1807 ATTGGATTCAAGCAGGCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTTGA 1866
 OY 101 SerThrThrGlyPheThrLeuTyrgLysLeuTyrgLysPheProHisAspLeuGlnProGly 120
 DB 1867 AGTACTACTGGAATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1926
 OY 121 LysLysTyrProThrValLeuPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 134
 DB 1927 AAGAAATATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1986
 OY 134 ----- 134

DB 1987 CCGTTTAAAGAGTCAAGATATTCGCTTGAAATACCTAGCCTCTAGATTATGTGTT 2046
 OY 134 ----- 134
 DB 2047 GTAGTATGACACACAGGGATCTCTGACCGAGGCTTAAATTTGAGGCCCTTTAAA 2106
 OY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
 DB 2107 TATTAATAGGTCATTAATGAAATGACATGACGATGAGGAAAGATCTCAATATCTAGCTCT 2166
 OY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyTyr 171
 DB 2167 CGATATGATTTTATGACTTATGATGATGATGATGATGATGATGATGATGATGATGATG 2226
 OY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 DB 2227 CTCTCCCTGATGACATTAATGACAGGTCAGATATCTTCAAGGTTGCTATTTCTGGGCC 2286
 OY 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 DB 2287 CCAGTCACCTGCTGGATCTTATGATACAGGATACAGGAACGTTATATGGGTACCCCT 2346
 OY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
 DB 2347 GACCAAGATGAAACAGGCTATTACTTAGATCTGGCCATGCAAGCAAGAAAGTTCCCC 2406
 OY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAla 251
 DB 2407 TTGAAACCAATCGTTTACTGCTCTTACATGATTTCTCGATGAGAAATGTCATTTTGA 2466
 OY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
 DB 2467 CATACAGATATATCTACTGATTTTATGAGAGGCTGGAAACCATATGATTTACAGATC 2526
 OY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291
 DB 2529 TATCTCTCGAGAGACACAGCATTAAGATCTCTGAAATGGGAGAACATTATGACATG 2586
 OY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 310
 DB 2587 CTTTTCACACTACCTTCAAGAAACCTTGATACGATATTGCTGCTTAAAGATGAT 2643

RESULT 4
 ABX12255
 ID ABX12255 strand; cDNA: 2649 BP.
 XX
 AC ABX12255;
 XX
 DT 19-MAY-2003 (first entry)
 XX
 DE cDNA encoding human serine protease HIPHUM46.
 XX
 KW Human; ss; gene; HIPHUM46; serine protease; gene therapy; osteoarthritis;
 KW serine protease activity modulation; dipyrilidyl peptidase activity;
 KW musculokeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KW Alzheimer's disease; paraneuronalcular palsy; Huntington's disease;
 KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
 KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KW colorectal tumour; ceroid lipofectinosis; allergic encephalomyelitis;
 KW multiple sclerosis; chromosome 19q21-q22.
 XX
 OS Homo sapiens.
 XX
 XX Key location/Qualifiers
 FH 1.2649
 FT CDS
 FT /tag= a
 FT /product= "HIPHUM46"
 FT /note= "Serine protease"
 XX
 PN GB2374869-A.
 XX

PD 30-OCT-2002.
 XX
 PF 22-JAN-2002; 2002GB-0001404.
 XX
 PR 23-JAN-2001; 2001GB-0001760.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Edbrooke MR, Lewis AP;
 XX WPI; 2003-150703/15.
 DR P-PSDB; AB007720.
 XX
 PT Identifying modulators of serine protease activity useful for treating
 PT musculoskeletal diseases, by contacting cell expressing a novel serine
 PT protease polypeptide with a compound and monitoring serine protease
 PT activity
 XX
 PS Claim 12; Page 22-26; 38pp; English.
 XX
 CC The invention relates to a method of identifying a substance that
 CC modulates serine protease activity, comprising contacting a cell such as
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
 CC virus, expressing a serine protease polypeptide (HIPHM 46), or its
 CC variant having dipeptidyl peptidase activity, or a serine protease
 CC isolated from the cell with a test substance and monitoring for serine
 CC protease activity. The method is useful for identifying a substance that
 CC modulates serine protease activity. A modulator of the serine protease is
 CC useful in the manufacture of a medicament for treatment or prophylaxis of
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 CC infection, Alzheimer's disease, paraneuronal nuclear palsy, myotonic
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
 CC Additional disease or amyotrophic lateral sclerosis.
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung
 CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 CC sclerosis. The present sequence represents cDNA encoding the human serine
 CC protease HIPHM46 which is located on chromosome 15q21-q22.
 XX
 SQ Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;
 Alignment Scores:
 Pred. No.: 8,4e-184 Length: 2649
 Score: 1645,50 Matches: 310
 Percent Similarity: 86,35% Conservative: 0
 Best Local Similarity: 86,35% Mismatches: 0
 Query Match: 97,95% Indels: 49
 DB: 25 Gaps: 1
 US-10-070-464-3 (1-310) x ABX12255 (1-2649)
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuValSerTyValAsn 20
 DB 1570 TTTGAAGGACACCAAGACTCCCTTTAGAGCATCACTGTAGTACGTAATAAT 1629
 QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysValIleSerGln 40
 DB 1630 CCTGAGAGAGTCAAGGCTGACTGACCGTGGCTCACTCATCTTTCGCGATCAGTCA 1689
 QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1690 CACTGTGACTCTTTATAGTAAGTATAGTACCGAAGAACCCACACTGTCTCCCTT 1749
 QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThr 80
 DB 1750 TACAGCATATCAAGGCTCGAAGATGACCACTTCCAAAGAACAAAGAAATTTGGGCCAC 1809
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
 DB 1810 ATTGTGATTGACGAGGCTCTTCTTCTGACTATACCTCTCCGAAATTTTCTTTTGA 1869
 QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120

DB 1870 AGTACTACGCAATTACATTGATGGATGCTTACAAAGCCCTCATGATATACAGCTGGA 1929
 QY 121 LysLysTyrProThrValLeuPheIleTyrGlyProGln----- 134
 DB 1930 AAGAAATATCCTACTGCTGCTGTTATATATGATGGTCCCTCAGGTGCAGTTGGTAATAT 1989
 QY 134 ----- 134
 DB 1990 CGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTTACGTTATGTGTT 2049
 QY 134 ----- 134
 DB 2050 GTAGTATAGAACAAAGGGATCTCTGACCGAGGGCTTAATTGAAGCGCCCTTAA 2109
 QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
 DB 2110 TATAAATGGGTCAATAGTAATTAATGATCAGTCAAGTCAAGTCAATATCTTACTTCT 2169
 QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyIleTyr 171
 DB 2170 CGATATGATTTCAATGACTTATGATGCTGTGGCATCCACGGCTGTCTTATGAGGATAC 2229
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 DB 2230 CTCTCCCTGATGGCATTTATGACAGAGTCAGATATCTTCAAGGTGCTATTTCTGGGCC 2289
 QY 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 DB 2290 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACAGGAACGTTATATGGGTCACTCT 2349
 QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuLysSerValAlaMetGlnAlaGluLysPhePro 231
 DB 2350 GACCAAGATGAACAGGGCTATTACTTACTAGATCTGTGGCCATGCAAGCAAAAGTTCCCC 2409
 QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
 DB 2410 TCTGACCAAAATCGTTTACTGCTCTTACATGTTTCCGTGATGAGAAATGTCATTTTGCA 2469
 QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
 DB 2470 CATACCAATATATTTCTGAGTTTATTTAGTGAGGCTGGAAGCCATATGATTACAGATC 2529
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291
 DB 2530 TATCTCTGAGAGACACAGCATTAAGATTCTGTAATCTGGAGAACATTATGAACTGCAT 2589
 QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 310
 DB 2590 CTTTTCGACTTACTTCAAGAAACCTTGGATCAGTATTTGCTCTTAAAGTGATTA 2646
 RESULT 5
 ABRK3322
 ID ABRK3322 standard; cDNA; 2671 BP.
 XX
 AC ABRK3322;
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPPIV related serine protease DPP-1.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 OS Homo sapiens.
 XX
 PN W0200231134-A2.

PD 18-APR-2002.
 XX 12-OCT-2001; 2001WO-US31874.
 XX 12-OCT-2000; 2000US-240117P.
 PR (FERR) FERRING BV.
 PA Qi S, Akinsanya KO, Riviere PJ, Junien J;
 PI WPI: 2002-444178/47.
 DR P-PSDB; ABG61591.
 XX
 XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 XX Claim 1; Page 53-54; 113pp; English.
 PS
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABR83322-ABR83343 encode human DPPR proteins.
 XX
 SQ Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other:
 Alignment Scores:
 Pred. No.: 8.5e-184 Length: 2671
 Score: 1645.50 Matches: 310
 Percent Similarity: 86.35% Conservative: 0
 Best Local Similarity: 86.35% Mismatches: 0
 Query Match: 97.95% Indels: 49
 DB: 24 Gaps: 1
 US-10-070-464-3 (1-310) x ABR83322 (1-2671)
 OY 1 PheGIUGLYThrLysAspSerProLeuGluHisIleuTYrValValSerTYrValAsn 20
 DB 1577 TTTGAAGGACACCAAGACTCCCTTAGAGCATCACCTGTAGTGAAGTAACTTAAT 1636
 OY 21 ProGIYluValAlThArGLeuThraSPARGLYTYrSerHisSerCysCysIleSerGln 40
 DB 1637 CCTGAGAGGTCACAAAGCTGACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1696
 OY 41 HisCysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1697 CACTGTGACTCTTCTTATAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 1756
 OY 61 TYrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrLysThr 80
 DB 1757 TACAAGCATCAAGCTCGAAGATGACCACTTGCACAAACAAAGAAATTTGGGCCACC 1816
 OY 81 IleuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePheSerPheGlu 100
 DB 1817 ATTTTGGATTTCAGAGTCTCTTCTGACTATATCTCTCCGAGAAATTTCTTTGAA 1876
 OY 101 SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGly 120
 DB 1877 AGTACTACTGATTTACATTTGATGGATGCTCTCAAGCCCTCAGATCTACAGCTCGA 1936
 OY 121 LysLysTYrProThrValLeuPheIleTYrGlyGlyProGln----- 134

DB 1937 AAGAAATATCTACTGTCGTGTTGATATANGTGAGTCTCAGGTGACGTGTAATAT 1996
 OY 134 ----- 134
 DB 1997 CGATTTAAAGATCAAGTATTTCCGCTTGAATACCTAGCCTCTTAGTTATGTGTT 2056
 OY 134 ----- 134
 DB 2057 GTAGGATAGACAAAGGGATCTCTGACCCAGAGGCTTAATTGAAAGGCCCTTTAA 2116
 OY 135 -----GlyGlnIleGluIleAspAspGlnValGlyGlyLeuGlnTYrLeuAlaSer 151
 DB 2117 TATAAATGGCTCAAAATGAAATTCAGATCAGGTGAGGACCTCAATATCTAGCTTCT 2176
 OY 152 ArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrPheSerTYrGlyTYr 171
 DB 2177 CGATATGATTTCACTTACCTTATGATGCTGTGGCATTCACGGCTGCTCTATGAGATAC 2236
 OY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 DB 2237 CTCTCCCTGATGGCATTAAATGACAGAGTCAGATATCTTCAAGGTGCTATGCTGGGCC 2296
 OY 192 ProValThrLeuThrPheThrPheTYrAspThrGlyTYrThrGluArgTYrMetGlyHisPro 211
 DB 2297 CCAATCACTCTGTGATCTCTTATGATACAGAAATACAGAAACCTTATATGAGGTACCT 2356
 OY 212 AspGlnAsnGlnGlnLYTYrTYrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
 DB 2357 GACCAAGATGACAGGGCTATTAATCTAGATCTGTGGCATGCAAGCAAGAAAGTTCC 2416
 OY 232 SerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
 DB 2417 TCTGACCAATTCGTTATCTGCTCTTACATGCTTCCGTGATGAGAAATGTCATTTGCA 2476
 OY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAspLeuGlnIle 271
 DB 2477 CATACAGTATATTAATGAGTTTGTGTGAGGCTGAAAGCCATATGATTTACAGATC 2536
 OY 272 TYrProGlnIuArgHisSerIleArgValProGluSerGlyGluHisTYrGlyLysHis 291
 DB 2537 TATCTCGAGGAGACACAGCATTAAGACTTCTGATATGGAGAACTTATGAACTGAT 2596
 OY 292 LeuLeuHisTYrLeuGlnGluAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310
 DB 2597 CTTTGGACACTACTTCAAGAAACCTTGATCAGATGCTGCTCTTAAAGTGATA 2653
 RESULT 6
 ABRK12892
 ID ABRK12892 standard; cDNA; 3106 BP.
 XX
 AC ABRK12892;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Human protease PRTS-9 cDNA sequence.
 KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
 KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KW cell proliferative disorder; developmental disorder; epilepsy;
 KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KW reproductive disorder; endometriosis; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 203..2851
 FT /*tag= a
 FT /product= "Human protease PRTS-9"
 PN MO200198468-A2.
 XX 27-DEC-2001.
 PD

```

XX 13-JUN-2001; 2001WO-US19178.
PF 16-JUN-2000; 2000US-212336P.
XX 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
XX 14-JUL-2000; 2000US-218946P.
XX
RA (INCY-) INCYTE GENOMICS INC.
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM,
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walla NR, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX WPI; 2002-090437/12.
DR P-PSDB; AAU74749.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 5; Page 166-167; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present nucleic acid sequence encodes
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 3106 BP, 928 A, 633 C, 704 G, 841 T, 0 other;
XX
Alignment Scores:
Pred. No.: 1,07e-183 Length: 3106
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
DB: 24 Gaps: 1
US-10-070-464-3 (1-310) x ABK12892 (1-3106)
OY 1 PHEGLUGLYTHLYRAHSPSERPROLEUGLHSHLEUTYRVAIVASERTYRVAIAEN 20
DB 1772 TTGGAAGGACCAAGACTCCCTTTAGAGCAATCACTGACTAGTACGTAAGTAAT 1831
OY 21 PROGLVGLVALTHARGLEUTHRAAPARGLYTYRSEHISERTYRSEGLIESERGIN 40
DB 1832 CCGTAGAGAGTACCAAGGCTGACTGACCGTGGCTACTCATCTTGCTGCTGACGTGACG 1891
OY 41 HISCYASPPHEPHEILESELYSYRSEAAAGLNLVAAAPROHIAVCYVALSERLEU 60
DB 1892 CACTGTGACTTTTATTAAGTAAAGTATATTAACCAAGAAATCCACACTGTGTGCTCTT 1951
OY 61 TYRVLVLEUSERSETPROGLUAPAPAPROTHRCYALVETHLYGLUBHETTPALATHR 80
DB 1952 TACAAGCTATCAAGCTCCGTAAGATGACCAACTTGCAAAAACAAGAAATTTGGCCACCC 2011
OY 81 ILEUAASPERALGLYPROLEUPROASPTYRTHPRPROGLULIEPHESERPHEGLU 100
DB 2012 ATTTCGATTTCAGCAGGCTCTCTCTGACTAATCTCTCCAGAAATTTTCCTTTGAA 2071

```

```

OY 101 SetThrThrclyPheThrleuTYRGLYMeLeuTYRlySPROHISAAPLEUGINPROGLY 120
DB 2072 AGTACTAGTGGATTTTACTGATTTGAGATCTCTACAGGCTCATGATCTACAGCTCGA 2131
OY 121 LVALYETYPROTHRVALLEUPHEILETYRGLYGLYPROGLN----- 134
DB 2132 AAGAAATATCTACTGTGTGCTGATATATAGTGTCTCTCAGGTGACGTTGTCATAATAT 2191
OY 134 ----- 134
DB 2192 CGGTTTAAAGAGTCAAGATTTCCGCTTGAATACCTTACCTCTAGGTATGTGTT 2251
OY 134 ----- 134
DB 2252 GTAGTATAGACAACAGGGGATCTCTCACCGAGGCTTAAATTTGAAGCGCTTTAA 2311
OY 135 -----GLYGLNLEGLUILEAPAPGINVALGLUJLYLEUGINTYRLEUAIAER 151
DB 2312 TATTAATGGGTCAATATAGAAATTTGACGATCAGGTGGAAGGACTCCAAATATCTAGCTTCT 2371
OY 192 ARGTYRASPHEILEAPLEUAPARGVALGLYLHISGLYTRPSEITYRGLYGLYTYR 171
DB 2372 CGATATGATTTCACTTGAATCTTGAATCGTGGGCGATCCAGGCTGTCTATGAGAGATAC 2431
OY 172 LEUSERLEUMERVALLEUMERGINARGSERAPDILEPHEARGVALAILEALAGLYALA 191
DB 2432 CTCTCCCTATGGCATTTAATGACAGGTCAAGATATCTTACGGGTCTATGCTGGGGCC 2491
OY 192 PROVALTHRLVETPILPHEPETHYRASPETHGLYTYRTHRCYUARGTYRMEGLYHISPRO 211
DB 2492 CAGGTACTCTGTGATCTTCTATATGATACAGGATACAGGAACGTTATATGGGTCAACCT 2551
OY 212 ASPPGINASRGINGLYTYRTRYLEUGLYSERVALIAMEGINLAAGLYSPHEPRO 231
DB 2552 GACCAAAAGAACAGGGCTATTACTTGAAGATCTGTGGCCATGCACAGAAAGTTCCCC 2611
OY 232 SERGLIPROASARGLEULEULEUENHISGLYPHELEUASGLUENVALHISPEALIA 251
DB 2612 TCTGAACCAAAATCGTTTACTGCTCTTACATGGTTTCTCGATGAGAAATGTCATTGGA 2671
OY 252 HISTRSERILEULEUSERPHELEUVALARGALAGLYLSPROTYRASPHEUGINILE 271
DB 2672 CATACAGATATATTACTGATGTTTATAGAGGGCTGAAAGCATATGATTTACGATC 2731
OY 272 TYRPROGLINGLUARGHISERTILEARGVALPROGLUSERGLYGLUHSERYGLUENHIS 291
DB 2732 TATCTCCAGAGACACACAGCATTAAGAGTTCCGTAATCCGAGACATTATGAATGCAAT 2791
OY 292 LEULEUHSERYRLEUGLINGLUENLEUGLYSERARGILEALALALEUYSVALIIE 310
DB 2792 CTTTTCGACTTCTTCAAGAAACCTTGGATCACGATTTGCTGCTCTTAAAGTGATA 2848

```

RESULT 7
 AAC85694
 ID AAC85694 standard; cDNA; 3120 BP.
 XX
 AC AAC85694;
 XX
 DT 29-JUN-2001 (first entry)
 XX
 DE Nucleotide sequence of human DPP8.
 XX
 KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
 KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
 KW growth hormone deficiency; glucose level; mucosal regeneration;
 KW non-insulin dependent diabetes mellitus; glucose intolerance;
 KW immunosuppression; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 214..2862
 FT CDS
 FT /tag= a

/product= "Human DPP8"

MO200119866-A1.

22-MAR-2001.

11-SEP-2000; 2000MO-AU01085.

10-SEP-1999; 99AU-0002762.

18-FEB-2000; 2000AU-0005709.

(UNSY) UNIV SYDNEY.

Abbot CA, Gorell MD;

WPI: 2001-281520/29.

P-PSDB; AAB87187.

New human dipeptidyl aminopeptidase (DPP8) useful for cleaving substrates, identifying inhibitors of DPP8 catalytic activity which have therapeutic uses, and for detecting activated T cells

Claim 16; Fig 2; 78pp; English.

This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable of hydrolyzing the peptide bond C-terminal to Pro in each of these compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for cleaving a substrate, and for detecting an activated T cell which involves measuring the level of DPP8 gene expression in a T cell. The level of DPP8 expression is detected by detecting the amount of DPP8 RNA in the cell. It is also useful for identifying a molecule capable of inhibiting the cleavage of the substrate by DPP8. Molecules identified as inhibiting DPP8 catalytic activity may be useful for treating diarrhoea, growth hormone deficiency, lowering glucose levels in non-insulin dependent diabetes mellitus and other disorders CC involving glucose intolerance, enhancing mucosal regeneration and as immunosuppressants.

Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:

Pred. No.: 1,08e-183 Length: 3120

Score: 1645.50 Matches: 310

Percent Similarity: 86.35% Conservative: 0

Best Local Similarity: 86.35% Mismatches: 0

Query Match: 97.95% Indels: 49

DB: 22 Gaps: 1

US-10-070-464-3 (1-310) x AAC85694 (1-3120)

1 PheGluGlyThrIleuAspSerProLeuGluHisIleuTyrValIleSerTyrValAsn 20

1783 TTTGAAGGACCAAAAGACTCCCTTTAGAGCATCAGTACGTTACGTAAT 1842

21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysSilesGln 40

1843 CCGAGAGAGGTGACAAAGGCTGACGCGTGTACTCATCTTGTGTCATCACTG 1902

41 HisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsnProHisCysValSerIleu 60

1903 CACTGTGACTCTTTATAGTAAGTAAAGTAACCAAGAAATCAACATGTGTCCCTT 1962

61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalaThr 80

1963 TACAGACTTCAAGTCTCTGAAGATGACCAACTTGCAAAACAAAGAAATTTTGCGCCAC 2022

81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100

2023 ATTGTGATTACAGACAGTCTCTTCTGTGACTATATCTCCAGAAATTTTCTTTTGA 2082

0

Qy 101 SerThrThrGlyPheThrIleuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120

Db 2083 AGTACTACTGATTTATTCATTTGATGGATGCTCTTACAAAGCTCATGATCTACAGCCTGGA 2142

Qy 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134

Db 2143 AAGAAATATCTTACTGTGCTTATATATATGATGTGCTCAGGTGAGTGTGAATAAT 2202

Qy 134 ----- 134

Db 2203 CCGTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTAGATTATGTGTT 2262

Qy 134 ----- 134

Db 2263 GTAGTGATAGCAACAGGGATCTGTGACCGAGGCTTAAATTGAAAGCGCTTTAA 2322

Qy 135 -----GlyGlnIleGluIleAspArgGlnValGluGlyLeuGlnTyrLeuAlaSer 151

Db 2323 TATAAATGGGTCAATATGAAATTCGATCAGTGGAGAGACTCCAAATATCTAGCTTCT 2382

Qy 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleGlyTyrSerTyrGlyGlyTyr 171

Db 2383 CGATATGATTTCTTACTTGAATCTGTGCGCATCCAGCGCTGCTCTATGAGATATC 2442

Qy 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191

Db 2443 CTCTCCCTGATGAGCTTAAATGACAGAGTCAGATATCTTACAGGTTGCTATTTGCTGGGCC 2502

Qy 192 ProValThrLeuThrIlePheTyrAspThrArgTyrThrGluArgTyrMetGlyHisPro 211

Db 2503 CCAGTCACCTCTGTGATCTTCTATATATACAGATACAGAGAAAGCTTATATGGCTCACCT 2562

Qy 212 AspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231

Db 2563 GACCGAAAGAACAGGCGCTTATCTTATGATCTGTGGCCATCGAACAAAGAAAGTTCCCC 2622

Qy 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251

Db 2623 TCTGAACCAATCGTTTACTGCTTACATGCTTCTCGATGAGAGAAATGCTCATTTTGCA 2682

Qy 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271

Db 2683 CATACCAATATATCTGAGTTTATGAGAGGCTGAGAAACCATATGATTTTACAGATC 2742

Qy 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyHisTyrGluLeuHis 291

Db 2743 TATCTCAGAGAGACACAGCATTAAGAGTCTCTGAATCGGAGAACATTTATGACATGCA 2802

Qy 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgGlyIleAlaLeuLysValIle 310

Db 2803 CTTTGTGACTACTTCAAGAAACCTTGGATCAGCATATGCTGTCTTAAAGTGATA 2859

RESULT 8

AAD38956

ID AAD38956 Standard; cDNA; 3120 BP.

AC AAD38956;

XX

DT 23-SEP-2002 (first entry)

XX

DE Human dipeptidyl peptidase 8 (DPP8) cDNA.

XX

KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis; autoimmunity; human immuno deficiency virus; HIV infection; cytotoxic; graft rejection; antidiabetic; antiinflammatory; immunosuppressive; antiviral; enzyme; gene; ss.

XX

OS Homo sapiens.

XX

Key Location/Qualifiers

FT CDS 214..2862

FT /tag= a

FT 0

/product= "Human DPP8 protein"

```

XX PN WO200234900-A1.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-AU01388.
XX PR 27-OCT-2000; 2000AU-0001078.
XX (UNSY ) UNIV SYDNEY.
XX Abbot CA, Gorrell MD;
XX MPI: 2002-454646/48.
XX P-PSDB; AAE24170.
XX
XX New dipeptidyl peptidase (DPP) peptidases, useful for screening
XX inhibitors of DPP catalytic activity, which may be employed to treat
XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV infection -
XX
XX Example: Fig 1; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptidases are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immuno deficiency virus) infection. The present
XX sequence is human DPP8 cDNA.
XX
XX Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,08e-183 Length: 3120
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
DB: 24 Gaps: 1
US-10-070-464-3 (1-310) x AAD38956 (1-3120)
QY 1 PheGluGlyThrThyAspSerProLeuGluHisLeuThrValValSerTyValAen 20
DB 1783 TTTGAAGGACCAAGACTCCCTTTAGGCATCACTGTAGTGTGCTGTTACGTAAT 1842
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTySerHisSerCysValIleSerGln 40
DB 1843 CCTGAGAGGTGACAAAGCTGACTGACCGTGTACTCATCTTGTGTGATCATGCTCG 1902
QY 41 HisCysAspPhePheIleSerTySerAsnGlnLysAsnProHisCysValIleSerLeu 60
DB 1903 CACTGTGACTTTTATTAAGTAAGTATTAACCAAGAAATCCACACTGTGTGCTCTT 1962
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 80
DB 1963 TACAAGCATATCAAGCTCTGAATGACCAATTGCAAAACAAAGAAATTTGGCCACAC 2022
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProGluIlePheSerPheGlu 100
DB 2023 ATTTTGATTTCAGCAGGTCTCTCTGCTGACTTACCTCCAGAAATTTTCTTTTGA 2082
QY 101 SerThrThrGlyPheThrLeuThrGlyMetLeuThrLysPheProHisAspLeuGlnProGly 120
DB 2083 AGTACTACTGAGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2142
QY 121 LysLysTyProThrValLeuPheIleTyArgGlyProGln----- 134
DB 2143 AAGAAATATCTACTGTCTGTTCATATATGTGTGTCTCTCAGGTGCAAGTGTGTAATAT 2202
QY 134 ----- 134
DB 2203 CGGTTTAAAGAGTACAGTATTTCCGCTTGAAATACCTAGCCTCTAGGTTATGTGTT 2262

```

```

QY 134 ----- 134
DB 2263 GTAGTATGACAAACAGGGATCTGTGACCGAGGCTTAATTTGAAGCGCCTTTAA 2322
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSer 151
DB 2323 TATTAATGGGTCAATATAGAAATTTGACGATCAGGTGGAAGGACTCCAAATATCTAGCTTCT 2382
QY 162 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyArgGlyTy 171
DB 2383 CGATATGATTTTATGATCTTAGATCTGTGGGATCAGCGGTGCTATGGAGGATAC 2442
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 2443 CTCTCCCGATGAGCTTAATGACAGGTCAGATATCTTACGAGGTGCTATGTGTGGGGCC 2502
QY 192 ProValThrLeuThrIlePheTyAspThrGlyTyThrGluArgTyMetGlyHisPro 211
DB 2503 CCAATCAGCTGTGATCTTCTATGATACAGATACAGGAAACGTTATATGGGTACCT 2562
QY 212 AspGlnAsnGluGlnGlyTyThrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
DB 2563 GACCAAGATGAACAGGCTATTAAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCT 2622
QY 232 SerGluProAsnArgLeuLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB 2623 TCTGAACCAATCGTTTACTGTCTTACATGAGTTTCTCGATGAGAAATGCAATTTTGA 2682
QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnIle 271
DB 2683 CATACCAATATTTACTAGATTTTATAGAGGCTGGAACCATATATTTACAGATC 2742
QY 272 TyrProGluGluArgHisSerIleArgValProGluSerGlyGluHisTyGluLeuHis 291
DB 2743 TATCTCTGAGAGACACAGCATTAAGTCTGTAATGGGAAACATTATGACATGAT 2802
QY 292 LeuLeuHisTyLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLysValIle 310
DB 2803 CTTTTCAGTACTTCAAGAAACCTTGATCAAGATATTCGCTTAAAGTGATA 2859
RESULT 9
AAH99934 standard, cDNA; 3143 BP.
ID AAH99934 standard, cDNA; 3143 BP.
AC AAH99934;
XX
XX 12-APR-2002 {first entry}
DT
XX
DE cDNA encoding 21953 human prollyl oligopeptidase.
XX
KW 21953 prollyl oligopeptidase; human; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antiarthritic; antiasclerotic; antiinflammatory;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 229..2877
XX FT /*tag= a
XX FT /product= "21953 prollyl oligopeptidase"
XX FT /note= "This region is specifically claimed in
XX claim 2"
XX
XX MO200179473-A2.
XX
XX 25-OCT-2001.
XX
XX 11-APR-2001; 2001WO-US40483.
XX

```


PR 18-APR-2000; 2000US-197508P.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Meyers RA, Williamson M;
 XX
 DR WPI; 2002-034353/04.
 XX P-PSDB; AAG78415.
 PT New polypeptides 21953, member of human prolyl oligopeptidase family,
 PT useful as diagnostic targets and therapeutic agents for controlling
 PT cancer, lymphoma and leukemia -
 XX
 PS Claim 7, Page 100-102; 121pp; English.
 CC This invention relates to an isolated 21953 human prolyl
 CC oligopeptidase, which is cytosolic, antidiabetic, antihypertensive,
 CC neuroprotective, antithyroid, dermatological, antipsoriatic,
 CC antidiabetic, ophthalmological, antiinflammatory, nootropic,
 CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
 CC antianxiety, cardiant, antithrombotic, anorectic and
 CC metabolic in its action. Uses include gene therapy, expression or
 CC activity of 21953 protein modulator, it is useful for identifying a
 CC compound which binds to it and can be used in preventing, treating
 CC or detecting a cellular proliferative or differentiative disorder.
 CC The 21953 molecules can act as novel diagnostic targets and therapeutic
 CC agents for controlling disorders associated with the aberrant activity
 CC or degradation of peptide hormones e.g., disorders associated with cell
 CC differentiation and proliferation such as cancer, immune function,
 CC reproductive, neurological and cardiovascular function. The 21953
 CC molecules are thus useful for treating and preventing cellular
 CC proliferative and differentiative disorders, haematopoietic neoplastic
 CC disorders, immune disorders such as autoimmune diseases, diabetes
 CC mellitus, arthritis, multiple sclerosis, asthma, Graves' disease,
 CC neuronal disorder, demyelinating diseases, vascular disorders and
 CC metabolism or pain disorders. This sequence represents the cDNA
 CC encoding sequence of 21953 human prolyl oligopeptidase.
 XX
 SQ Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 1, 09e-183 Length: 3143
 Score: 1645.50 Matches: 310
 Percent Similarity: 86.358 Conservative: 0
 Best Local Similarity: 86.358 Mismatches: 0
 Query Match: 97.958 Indels: 49
 DB: 24 Gaps: 1
 US-10-070-464-3 (1-310) X AAH99934 (1-3143)

QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
 DB 2158 AAGAAATATCTTACTGTGCTGTTATATATGTTGGTCCCTCAGTGCAGTTGGAAATAAT 2217
 QY 134 ----- 134
 DB 2218 CGGTTTAAAGATCAATATTTCCGCTTGAATACCTAGCCTCTTACGTTATGCTT 2277
 QY 134 ----- 134
 DB 2278 GTATGATATGACACAGGGGATCTCTGACCGAGGCTTAAATTGAAGCGCCTTTAA 2337
 QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
 DB 2338 TATTAATATGGGTCAATATGAAATTCATGATCAGTGGAAAGCATCTCAATATCTTACTTCT 2397
 QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171
 DB 2398 CGATATGATTTCTATTTGACTTATGATGCTGTGGGCATCCAGCGCTGTCTTATGAGGATAC 2457
 QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 DB 2458 CTCTCCCTGATGGCATTTAATGACAGGTCAGATATCTTCAGGCTTGCTATTTGCTGGAGCC 2517
 QY 192 ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 DB 2518 CCAGTCACTGCTGGATCTTATGATATCAGGATACAGGAACGTTATATGGGTACCCCT 2577
 QY 212 ArgGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
 DB 2578 GACCGAATATGAACAGGGCTATTAATCTTATGATCTGTGGCCATCAAGCAAAAAAGTTCCCC 2637
 QY 233 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
 DB 2638 TTGTAACCAATCGTTTACTGCTTCACTGATGTTCTCGATGAGAAATGTCATTTTGCA 2697
 QY 252 HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
 DB 2698 CATACCAATATTAATCTGATGTTTATGAGAGGCTGGAAACCATATGATTTACAGATC 2757
 QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyrGluLeuHis 291
 DB 2758 TATCTCTCAGGAGAGACACAGCATTAAGATTTCTGATCGGGAACATTTATGAACGCAAT 2817
 QY 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaIleLeuValIle 310
 DB 2818 CTTTTCGACTTACTTCAAGAAAACTTGGATCAGGATTTGCTCTAAAAAGTATA 2874
 RESULT 10
 ABRK3331
 ID ABRK3331 standard; cDNA; 4676 BP.
 XX
 AC ABRK3331;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPRP-1 splice variant #7.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinnesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX

PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 DR MPI: 2002-444178/47.
 DR P-PSDB; ABG61500.
 PR New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 PS Disclosure: Page 72-73; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcer, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
 XX
 SO Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,616-183 Length: 4676
 Score: 1644.50 Matches: 308
 Percent Similarity: 99.35% Conservative: 0
 Best Local Similarity: 99.35% Mismatches: 1
 Query Match: 97.89% Indels: 2
 DB: 24 Gaps: 1
 US-10-070-464-3 (1-310) x ABK83331 (1-4676)
 QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuValValSerTyValAsn 20
 DB 1783 TTTGAAGCACCAAGACTCCCTTTAGGACATCACTGTAGTACGTTACGTAAT 1842
 QY 21 ProGlyValValThrArgLeuThrAspArgGlyTyTySerHisSerCysValIleSerGln 40
 DB 1843 CCGAGAGGTACAAAGGCTGACCTGACCTGCTCACTTCTTGCTGACATCAGTCA 1902
 QY 41 HisCysAspPhePheIleSerLysTyTySerAsnGlnLysAsnProHisCysValIleSerLeu 60
 DB 1903 CACTGTGACTCTTTATAGTAAGTATAGTACCAAGAAATCCACAGTGTGCTCCCTT 1962
 QY 61 TyrLysLeuSerSerProGluAspAspProThrTyLysThrLysGluPheThrAlaThr 80
 DB 1963 TACAGAGCTATCAAGCTCGAAGATGACCCCACTTCAAAACAAAGAAATTTTGCCCA 2022
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyTyThrProProGluIlePheSerPheGlu 100
 DB 2023 ATTGTGATTCAGCAGGCTCTTCTTCTGACTATCTCTCCGAAATTTCTCTTTGAA 2082
 QY 101 SerThrThrGlyPheThrLysTyGlyMetLeuTyTyLysProHisAspLeuGlnProGly 120
 DB 2083 AGTACTACTGAGATTACATTGATGAGATGCTCTCAAGCCCTACGATCTACAGCCTGGA 2142
 QY 121 LysLysTyTyProThrValLeuPheIleTyTyGlyGlyProGlnGlyGlnIleGluIleAsp 140
 DB 2143 AAGAAATATCCACTGTGCTGTTCATATATGATGATGCTG---GTCAAATAGAAATTTAC 2198

QY 141 AspGlnValGluGlyLeuGlnIleTyrLeuAlaSerArgTyTyAspPheIleAspLeuAspArg 160
 DB 2199 GATCAGGTGGAGAGACTCCAAATATCTAGCTTTCCATATGATTTCACTTACATGATCGT 2258
 QY 161 ValGlyIleHisGlyTyTySerTyTyGlyGlyTyTyLeuSerLeuMetAlaLeuMetGlnArg 180
 DB 2259 GTGGGCATCCACGGCTGGTCCATGAGAGATACCTCTCCCTGATGSCATTATTCGAGAG 2318
 QY 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyTyAsp 200
 DB 2319 TCAGATATCTTCAGAGGTTCCTATGCTGGGCCCCAGATCACTCTGGGATCTTATGAT 2378
 QY 201 ThrGlyTyTyThrGluArgTyTyMetGlyHisProAspGlnAsnGlnGlnGlyTyTyLeu 220
 DB 2379 ACGAGATACCGAAAGCTTATATGGGTCCCTGACCAAGATGAACAGGGCTATTACTTA 2438
 QY 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240
 DB 2439 GGATCTGTGGCCATCCAAAGCAAGAAAGTTCCCTTGAAACCAATCGTTTACTGCTCTTA 2498
 QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260
 DB 2499 CATGCTTTCCTGGATGAGAAATGTCATTTTGCACATACCAATATTACTGAGTTTCTTA 2558
 QY 261 ValArgAlaGlyLysProTyTyAspLeuGlnIleTyTyProGlnGluAsnHisSerIleArg 280
 DB 2559 GTGAGGGCTGGAAAGCCATATGATTTACAGATCTATCTCCAGAGAGACACAGCATMAA 2618
 QY 281 ValProGluSerGlyGluHisTyTyGluLeuHisIleLeuHisTyTyLeuGlnGluAsnLeu 300
 DB 2619 GTTCTGATAGGAGAAACATTATGATGATCGATCTTTTGGCACTTACTTCAAAAACCTT 2678
 QY 301 GlySerArgIleAlaIleLeuLysValIle 310
 DB 2679 GGATCAGCATTTGCTGCTTAAGTATA 2708
 RESULT 11
 ID ABK83328 standard; cDNA, 4309 BP.
 XX
 AC ABK83328;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPPR-1 splice variant #4.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 OS Homo sapiens.
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 DR MPI: 2002-444178/47.
 DR P-PSDB; ABG61597.
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPP-IV proteins.

XX Sequence 4829 BP, 1466 A; 886 C; 1017 G; 1460 T; 0 other;

Alignment Scores:

Pred. No.:	4,84e-181	Length:	4829
Score:	1625.50	Matches:	310
Percent Similarity:	85.87%	Conservative:	0
Best Local Similarity:	85.87%	Mismatches:	0
Query Match:	96.76%	Indels:	51
DB:	24	Gaps:	1

US-10-070-464-3 (1-310) x ABK83327 (1-4829)

QY 1 PhegluGlyThrIlyAspSerProLeuGluHsiSleuTYrValValSerTYrValAsn 20
 DB 1783 TTTGAGGACCAAGAGCTCCCTTTAGAGCATCCGTGAGTACGATGATTAAT 1842
 QY 21 ProgluValuValThrArgLeuThrAspArgIlyTyrSerHisSerCysGluSerGln 40
 DB 1843 CCTGAGAGGTGACAGGCTTACTGACCGTGGCTTCTACATCTTGTGCTGACGTAG 1902
 QY 41 HisCysAspPhePheIleSerIlySerAsnGlnIlyAspProHisCysValSerLeu 60
 DB 1903 CACGTGACTCTTTATATAGTAGTAGTAACCAAGAAATCCACATGTGTGCCCTT 1962
 QY 61 TyrllysleuSerSerProGluAspAspProThrCysIlyThrIlyGluPheTrpAlaThr 80
 DB 1963 TACAGAGCTATCAAGTCCCTGGAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACC 2022
 QY 81 IleleuAspSerAlaGlyProleuProAspIlyThrProProGluIlePheSerPheGlu 100
 DB 2023 ATTGTGATTCAGCAGGTCCTTCTCTCTACTACTACTCTCTCAAGAAATTTCTCTTTTAA 2082
 QY 101 SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrIlyProHisAspLeuGlnProGly 120
 DB 2083 AGTACTACTGATTTACTATGTATGGAGTGTCTACAAAGCCCTCATGATCTACAGCTGGA 2142
 QY 121 LysIlySerProThrValleuPheIleTYrGlyIly--ProGln----- 134
 DB 2143 AAGAAATATCTTACTGTGCTGTTCATATATGTGTCTCTCAGGTGACGTTGTGAATA 2202
 QY 134 ----- 134
 DB 2203 ATCGGTTAAAGAGTCAAGATTTCCGCTTGAATACCTTACCTCTAGGTTATGCG 2262
 QY 134 ----- 134
 DB 2263 TTGTAGTATAGACAAACAGGGATCCGTGACCGAGGCTTAATTGAAAGCGCCTTTA 2322
 QY 135 -----GlyGlnIleGluIleAspArgGlnValGluGlyLeuGlnIlyrleuAla 151
 DB 2323 AATATAAATGGTCAATATGAAATGACATGAGTGGAGGACTCCAAATATCTAGCTT 2382
 QY 151 eArqTYrAspPheIleAspLeuAspArgValIleHisGlyTrpSerTYrGlyIlyT 171
 DB 2383 CTCGATATGATTTCTATGATCTGATCTGTGGGATCCAGGCTGTGCTTATGAGGAT 2442
 QY 171 YrleuSerleuMetAlaIleuMetGlnArgSerAspIlePheArgValAlaIleAlaIly 191

DB 2443 ACCTCTCCGTGATGATTAATGACAGGTGACAGATATCTTCAGGGTCTATTGCTGGG 2502
 QY 191 IaProValThrLeuTrpIlePheTYrAspThrGlyTYrThrGluArgTYrMetGlyHisP 211
 DB 2503 CCCAGTACCTGTGGATCTTCTATGATGATGATACAGGAAAGTTATATGAGGTGACC 2562
 QY 211 roAspGluAspGlnGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGluIlyPheP 231
 DB 2563 CTGACCAAGATGACAGGCTTATTAAGATCTGTGCTGCTGCAATGCAAGCAAAAGTTCC 2622
 QY 231 roSerGluProAspArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251
 DB 2623 CCGTGAACCAATGCTTATGCTGCTTACATGCTTCTGATGATGATGATGATGATGATG 2682
 QY 251 IaHisThrSerIleLeuLeuSerPheLeuValAlaGlyIlySerProTYrAspLeuGln 271
 DB 2683 CACATACCAAGTATATTTACTGAGTTTATGAGGCTGGAAGCAATATGATTTTACAGA 2742
 QY 271 leTYrProGluGluArgHisSerIleArgValProGluSerGlyGluHisIlyTYrGluLeuH 291
 DB 2743 TCTATCTCAGGAGAGACACAGCATTAAGATTCTGAAATCCGAGAAACATTATGAACTGC 2802
 QY 291 IsleuLeuHisIlyrleuGlnGluAsnLeuGlySerArgIleAlaIleLeuIlyValIle 310
 DB 2803 ATCTTTGACATACCTTCAAGAAAACCTTGATGACGTAATGCTGCTCTAAAAAGTGATA 2861
 RESURF 13 0 4
 AAH15009 standard; cDNA; 2161 BP.
 ID AAH15009;
 AC AAH15009;
 DT 26-JUN-2001 (first entry)
 XX Human cDNA sequence SEQ ID NO:12963.
 DE Human;
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 PN EPI074617-A2.
 XX 07-FEB-2001.
 PD 28-JUN-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0189767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isegaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Iishi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX Claim 8; SEQ ID 12963; 2537bp + CD ROM; English.
 PS The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH13632
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

SO Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;

Alignment Scores:

Pred. No.:	5.14e-155	Length:	2161
Score:	1401.00	Matches:	269
Percent Similarity:	88.05%	Conservative:	11
Best Local Similarity:	84.59%	Mismatches:	20
Query Match:	83.39%	Indels:	18
DB:	22	Gaps:	5

US-10-070-464-3 (1-310) x AAH15009 (1-2161)

```

OY      1 PheGluGluThrTyrAspSerProLeuGluHisLeuTyrValIleSerTyrValaen 20
DB      988 TTGAAAGGACCAAGACCTCCCTTTAGAGCATCCTTACGTAAGTCACTTACGTAAGT 1047
OY      21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
DB      1048 CCTGAGAGAGTGACAAAGCTGACGACGTAAGTCACTTCTGTCGATCAGTACG 1107
OY      41 HisCysAspPhePheLeuSerTyrSerAsnGlnLysAspProHisCysValSerLeu 60
DB      1108 CACTGAGACTTCTTATTAAGTAAAGTAACTCAACAAAGTCAACACGTCGTGCTCCCT 1167
OY      61 TyrLysLeuSerSerProGluAspAspProThrCysValThrLeuGluPheTrpAlaThr 80
DB      1168 TACAAAGCTTCAAGTCTGTAAGATGACCCAACTTGCAAAACAAGAAATTTGGGCGCAC 1227
OY      81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
DB      1228 ATTTGAGATTGACAGAGTCTCTCTCTGACTACTACTCCCAAGAAATTTCTCTTTGAA 1287
OY      101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
DB      1288 ACTACTACTGATTTTACATTTGATGAGATGCTTCAACAGCCCTCATGATCTTACAGCCTGGA 1347
OY      121 LysLysTyrProThrValLeuPheLeuTyrGlyGlyProGlnGlyGlnIleGluIleAsp 140
DB      1348 AAGAAATATCCTACTGCTGCTGTTCAATATGTTGTCCTCAGTGAGTTG---GTGAAAT 1404
OY      141 AspGlnValGlyGlyLeuGlnTyr-----LeuAlaSer---ArgTyrAsp 154
DB      1405 AATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTGAGCCCTCTAGATTATGAG 1464
OY      155 PheIleAspLeuAspArgValGlyIle-----HisGlyTyrSerTyrGlyGlyTyrLeu 172
DB      1465 GTTGATGATGATGACCAAGGAGATCCTGTCAACCGAGGCTTAAATTTAAAGCGGCTTT 1524
OY      173 SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
DB      1525 AATAATATAA-----ATGCTTGCTATTCCTGGGGCCCA 1557
OY      193 ValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAsp 212
DB      1558 GTCACCTGCTGATCTTCTATATACAGATACACGGAACGTTAATATGGGTACCCCTGAC 1617
OY      213 GlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaIleMetGlnAlaGlyLysPheProser 232
  
```

```

DB      1618 CGAATATGACAGGCGCTATTACTTAGATCTGTGGCCATGCAAGCAAGAAAGTCCCTCT 1677
OY      239 GlnPheAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHis 252
DB      1678 GAACCAATCTGTTCTGCTCTTACATGAGTTTCTCTGATGAGAAAGTCATTTTGCACAT 1737
OY      253 ThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyr 272
DB      1738 ACCAGATATTTACTGATTTTATGAGAGGCTGGAAGCCATATGATTACAGATCTAT 1797
OY      273 ProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHisLeu 292
DB      1798 CCTCAGAGAGACACAGCATTAAGATCTCTGATTCGGGAGAACATTATGAACTGCATCTT 1857
OY      293 LeuHisTyrLeuGlnGluAsnLeuGlySerArgGlyLysAlaIleAlaLeuLysValIle 310
DB      1858 TTGCACTACTCTTCAAGAAACCTTGATTCACGATATTGCTGCTTAAAGTGATTA 1911
  
```

RESULT 14

AA085696
 ID AAC085696 standard; cDNA; 1669 BP.

XX AAC085696;

XX 29-JUN-2001 (first entry)

XX Nucleotide sequence encoding human DPP8 318Thr-658Val+759Ala-882Ile.

XX Human; dipeptidyl aminopeptidase; DPP8; prollyl oligopeptidase;

XX dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;

XX growth hormone deficiency; glucose level; mucosal regeneration;

XX non-insulin dependent diabetes mellitus; glucose intolerance;

XX immunosuppression; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 2..1399

XX FT /tag= a

XX PN /product= "Human DPP8 318Thr-658Val+759Ala-882Ile"

XX MO200119866-A1.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000MO-AU01085.

XX 10-SEP-1999; 99AU-0002762.

XX 18-FEB-2000; 2000AU-0005709.

XX (UNSV) UNIV SYDNEY.

XX Abbot CA, Gorell MD;

XX WPI; 2001-281520/29.

XX P-PSDB; AAB47189.

XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving

XX PT substrates, identifying inhibitors of DPP8 catalytic activity which

XX have therapeutic uses, and for detecting activated T cells -

XX Claim 21; Page 72-74; 78pp; English.

XX The sequences given in AAC085696-97 encode fragments of human

XX dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for

XX H-Ala-Pro-PNA, H-Gly-Pro-PNA and H-Arg-Pro-PNA. Therefore, it is a

XX prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable

XX of hydrolysing the peptide bond C-terminal to pro in each of these

XX compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for

XX cleaving a substrate, and for detecting an activated T cell which

XX involves measuring the level of DPP8 gene expression in a T cell. The

XX level of DPP8 expression is detected by detecting the amount of DPP8

CC RNA in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.

XX Sequence 1669 BP; 524 A; 336 C; 336 G; 473 T; 0 other;

Alignment Scores:

Pred. No.:	2,09e-152	Length:	1669
Score:	1377.50	Matches:	259
Percent Similarity:	83.55%	Conservative:	0
Best Local Similarity:	83.55%	Mismatches:	0
Query Match:	81.99%	Indels:	51
DB:	22	Gaps:	1

US-10-070-464-3 (1-310) x AAC85696 (1-1669)

```

Oy 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValValSerTyrValaAn 20
Db 620 TTGAAAGGACCAAAAGACTCCCTTTAGAGCATCCTGAGTACGTACGTAACGTAAT 679
Oy 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
Db 680 CCTGGAGAGGTACAAAGCTGACTGACCGTGGCTACTCATCTTCTGGCATCAGTACG 739
Oy 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
Db 740 CACTGTGACTTTTATTAAGTAAGTAACTAACCAAGAAATCCACACGTGTGTCCTT 799
Oy 61 TyrLysLeuSerSerProGluAspAspProThrCysValSerThrLysGluPheThrAlaThr 80
Db 800 TACAAAGCTATCAAGTCTTAAGATGACCCAACTTGCAAAACAAAGAAATTTGGGCCACC 859
Oy 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGln 100
Db 860 ATTTGGATTGACGAGGTCTCTCTGCTACTATCTCCCAAGAAATTTCTCTTTTGA 919
Oy 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
Db 920 AGTACTACTGAGATTACATTTGATGGAGTCTCTCAACACCTCATGATCTTACAGCCTGGA 979
Oy 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnGlyIleGlnIleAsp 140
Db 980 AAGAAATATCTACTGCTGCTGCTCATATATGTTGGTCTCTCAG----- 1021
Oy 141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160
Db 1021 ----- 1021
Oy 161 ValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180
Db 1021 ----- 1021
Oy 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAsp 200
Db 1022 -----CTTCTATTTGCTGGGGCCCAAGTACTCTGTGATCTTCTATGAT 1066
Oy 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnIleTyrTyrLeu 220
Db 1067 ACAGATACACAGGACGTTATGATGGTCACTGACCAAGACAAATGAACAGGCGCTATTACTTA 1126
Oy 221 GlySerValAlaMetGlnAlaGluLysPheProSerGlnProAsnArgLeuLeuLeu 240
Db 1127 GGATCTGGGCCATCCAAAGCAAAAGTCCCTCTGAAACCAAACTGTTTACTGCTCTTA 1186
Oy 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260
Db 1187 CATGTTTCTTGAGAGAGATGTCATTGTGACATACCAAGTATTTACTGAGTTTCTTA 1246
Oy 261 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280

```

```

Db 1247 GTGAGGGCTGAAAGCAATATGATTATACGATCTATCTCAGAGAGACACACATTAAGA 1306
Oy 281 ValProGluSerGlyGluHisTyrGluLeuHisLeuHisTyrLeuGlnGluAsnLeu 300
Db 1307 GTTCTGAACTGGAGAAACATTAATGAACTGATCTTTTGCACATCACTTCAAGAAACCTT 1366
Oy 301 GlySerArgIleAlaLeuLeuValIle 310
Db 1367 GCATCAAGTATTCCTGCTCTTAAGATGATTA 1396

```

RESULT 15

ABNS9775
 ID ABNS9775 standard; cDNA; 2668 BP.

XX ABNS9775;

XX 28-JUN-2002 (first entry)

DE Novel human coding sequence SEQ ID NO: 186.

XX Human; anti-nausea; vulnerrary; anti-inflammatory; immunomodulator;

XX anti-fertility; cerebroprotective; cytoskeletal; rheumatic; gene therapy;

XX neuroprotective; anti-parkinsonian; protein therapy; EST;

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Weinman T, Drmanac RT;

DR WPI; 2002-292408/33.

XX P-PSDB; ABB97362.

PT An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis -

PS Claim 1; SEQ ID NO 186; 509bp; English.

XX The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat

CC stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a coding sequence of the

CC invention.

XX Sequence 2668 BP; 796 A; 564 C; 592 G; 716 T; 0 other;

Alignment Scores:

Pred. No.:	4,26e-152	Length:	2668
Score:	1377.50 <td>Matches:</td> <td>259 </td>	Matches:	259
Percent Similarity:	83.55% <td>Conservative:</td> <td>0 </td>	Conservative:	0
Best Local Similarity:	83.55% <td>Mismatches:</td> <td>0 </td>	Mismatches:	0
Query Match:	81.99% <td>Indels:</td> <td>51 </td>	Indels:	51
DB:	24	Gaps:	1

US-10-070-464-3 (1-310) x ABNS9775 (1-2668)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:29:02 ; Search time 48.5672 Seconds
(without alignments)
2817.309 Million cell updates/sec

Title: US-10-070-464-3
Perfect score: 1680
Sequence: 1 FEGTKDSPLEHMLVVSYYN.....HLHYIQENLGRIMALKVI 310

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-O=/cgn2_1/USPTO.spool/US10070464/rn1at 15102003 113554 24845/app query.fasta_1.2652
-DEL=frme+p2n.model -DEV=x1h
-DB=issued_patents_NA -OPMT=fastap -SUFFIX=nni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cai
-LIST=45 -DOCLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464 @CGN 1.1 146 @runat 15102003 113554 24845 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	755.5	45.0	823	4	US-09-280-116-171 Sequence 171, App
2	405.5	24.1	612	4	US-09-392-184-31 Sequence 31, Appl
3	337.5	20.1	3407	4	US-10-002-593-5 Sequence 5, Appl
4	333.5	19.9	2924	5	PCT-US93-07923-1 Sequence 1, Appl
5	337	19.5	5496	4	US-09-462-284-1 Sequence 1, Appl
6	302	18.0	2815	1	US-08-220-491A-1 Sequence 1, Appl
7	302	18.0	2815	1	US-08-619-280A-1 Sequence 1, Appl
8	302	18.0	2815	2	US-08-940-391-1 Sequence 1, Appl
9	267	15.9	4982	3	US-08-699-103B-1 Sequence 1, Appl
10	267	15.9	4982	4	US-09-229-059-1 Sequence 1, Appl
11	267	15.9	4982	4	US-09-628-133-1 Sequence 1, Appl
12	263	15.7	543	4	US-09-221-017B-253 Sequence 253, App

C	13	252	15.0	657	4	US-09-221-017B-646	Sequence 646, App
	14	210	12.5	535	4	US-09-389-681-428	Sequence 428, App
	15	210	12.5	535	4	US-09-620-405B-428	Sequence 428, App
	16	210	12.5	535	4	US-09-433-826B-428	Sequence 428, App
	17	210	12.5	535	4	US-09-604-287A-428	Sequence 428, App
	18	207	12.3	1896	3	US-09-016-080-2	Sequence 2, Appl
	19	193	11.5	502	4	US-09-280-116-172	Sequence 172, App
	20	159.5	9.5	3085	4	US-09-221-017B-1010	Sequence 1010, App
	21	143.5	8.5	1869	2	US-08-664-646A-1	Sequence 1, Appl
	22	143.5	8.5	1869	2	US-09-066-285-1	Sequence 1, Appl
	23	143.5	8.5	1869	3	US-09-261-006-1	Sequence 1, Appl
	24	143.5	8.5	1869	3	US-08-951-088-1	Sequence 1, Appl
	25	143.5	8.5	1869	4	US-09-609-566-1	Sequence 1, Appl
	26	143.5	8.5	1869	4	US-09-609-570-1	Sequence 1, Appl
	27	143.5	8.5	1869	4	US-09-427-372-1	Sequence 1, Appl
	28	143.5	8.5	1869	4	US-09-693-554-1	Sequence 1, Appl
	29	138	8.2	1845	4	US-09-252-991A-3489	Sequence 3489, App
	30	110.5	6.6	1230025	4	US-09-198-452A-1	Sequence 1, Appl
	31	110	6.5	904	4	US-09-221-017B-58	Sequence 58, Appl
	32	96.5	5.7	815	4	US-09-221-017B-962	Sequence 962, App
	33	95.5	5.7	1032	4	US-09-724-623-17	Sequence 17, Appl
	34	95	5.7	679	4	US-09-221-017B-13	Sequence 13, Appl
	35	94	5.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
	36	94	5.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
	37	92.5	5.5	12835	4	US-08-961-527-141	Sequence 141, App
	38	91.5	5.4	1710	3	US-09-068-960-14	Sequence 14, App
	39	91	5.4	3601	4	US-09-068-960-12	Sequence 2, Appl
	40	90.5	5.4	1389	4	US-09-134-001C-1463	Sequence 1463, App
	41	90	5.4	35081	2	US-08-752-760A-1	Sequence 1, Appl
	42	89.5	5.3	1830121	4	US-09-557-884-1	Sequence 1, Appl
	43	89.5	5.3	1830121	4	US-09-643-990A-1	Sequence 1, Appl
	44	89	5.3	2046	1	US-07-688-299-12	Sequence 12, Appl
	45	89	5.3	2046	1	US-07-980-517A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-280-116-171
Sequence 171, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280, 116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 171
LENGTH: 823
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Prolyl oligopeptidases
US-09-280-116-171

Alignment Scores:

Pred. No.: 6.56e-80 Length: 823
Score: 755.50 Matches: 164
Percent Similarity: 88.95% Conservative: 5
Best Local Similarity: 86.32% Mismatches: 5
Query Match: 44.97% Indels: 16
DB: 4 Gaps: 2

US-10-070-464-3 (1-310) x US-09-280-116-171 (1-823)

QY 136 Gmltlegluilaaapapgnlvalgluglyl-leuglntyrleualaserargtyraspph 155
Db 3 CAATGGAATTCAGCATGAGTGAAGACATCCATATCTTACGCTTCGATATATTT 62
QY 155 ellaaaplueapapavaglyllehlscltyrpserrygylgylytyrleuserleuwe 175
|||||

Db 63 CATTGACCTAGATGCTGGGATCCAGCGCTGGCTTATGAGAGATACCTCTCCCTGAT 122
Qy 175 TAlaLeuMeGlnAArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrIle 195
Db 123 GGCATTATGAGAGGTCAGATCTTTCAGAGGTTCTTCTGGGTCCTCCAGTCACTCT 182
Qy 195 UTPIlePheTyraPThrGlyTyThrGluArgTyIleGlyIleHisProAspGlnAsnGly 215
Db 183 GTGATCTCTTCAATATACAGATACACGAAAGCTTATATGAGTCACTCCAGCAAGATGA 242
Qy 215 UGlnGlyTyTyThrLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAs 235
Db 243 ACAGAGGCTATTAATCTAGATCTGGCCATCAAGCAAGAAAGTTCCTCCCTGAAACCAA 302
Qy 235 nArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIle 255
Db 303 TCGTTACTGCTCTTCAATGATGTTCTCGATGAGAAATGCCATTTTGACATACACAGAT 362
Qy 255 eLeuLeuSerPheLeuValArgAlaGlyLysProTyraSpleuGln--IleTyr---- 272
Db 363 ATTAAGTGTATTTAGTGAAGGCTGGAAACAGATATTTACAGATCTTTATATTTT 422
Qy 273 -----ProGlnGluArgHis-SerIleArgVal-ProGluSer-Gl 285
Db 423 GTTGTGTGTAAGATCTATCTCTCAGGAGAGACACCGCATTAAGAGTTCTCGAAATCGGAG 482
Qy 285 YGluHisTyrgIleuHisLeuLeu-HisTyThrLeuGlnGlu-AsnLeuGlySer-ArgIle 304
Db 483 AGAACATTATTAACATGCATCTTTTGACACTTCAAGAAACCTTTGGATCACTTAT 542
Qy 304 eAlaAlaLeuLysVal 309
Db 543 TGCTGCTCCTAAATT 558
RESULT 2
US-09-392-184-31/C
; Sequence 31, Application US/09392184
; Patent No. 6395889
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: PROTEASE HOMOLOGS
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392,184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; LOCATION: (1)...(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31
Alignment Scores:
Pred. No.: 1,02e-38 Length: 612
Score: 405.50 Matches: 100
Percent Similarity: 67.07% Conservative: 10
Best Local Similarity: 60.98% Mismatches: 49
Query Match: 24.14% Indels: 7
DB: 4 Gaps: 1
US-10-070-464-3 (1-310) x US-09-392-184-31 (1-612)
Qy 127 LeuPheIleTyrgIleuHisLeuLeu-HisTyThrLeuGlnGlu-AsnLeuGlySer-ArgIle 146
Db 543 GTGCTNATCCGGGNAACN-----NAAAGATTGTGCGATCAGTGGAGGACTC 493

Qy 147 -GlnTyThrLeuAlaSerArgTyraSpleuIleAspLeu-AspArgValGlyIleHisGlyT 166
Db 492 CCATATCTTAATCTCTGAATATGATTTCAATGACNTAAAGATCGTGGCAATCCNCGCT 433
Qy 166 rPseTyrgIleGlyTyThrLeuSerLeuMetAlaLeuMetGlnAArgSerAspIlePheArgV 186
Db 432 G-TCTATGAGAGATACCTCTCCCTGATGCGCANTAAATACAGAGCTCAGATATCMTCAAGG 374
Qy 186 AlAlaIleAlaGlyAlaProValThrLeuPrlIlePheTyraPThrGlyTyThrGluA 206
Db 378 TTCT-ATTCTGGGGCCCACTGATCTGTGATCTTATGAACAGAAACNCGGAC 315
Qy 206 rGlyrMetGlyHisPheProAspGlnAsnGluGlnGlyTyThrLeuGlySerValAlaMetG 226
Db 314 GTTANATGGGTACACCTCCTCCTGAAACAGATGACAGGCTATTAAGATCTGTGGCAATC 255
Qy 226 lAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspG 246
Db 254 CANCAGGAAGTTCCTCCCTGAAACCAATNNTTCTTANATGTTTCTCGATT 195
Qy 246 lueAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysP 266
Db 194 AGGATTTCCANTTTTCAACAAACAGNANATTAATTAAGTTTATGTAAGGCTGGAAAC 135
Qy 266 rOTyraSpleuGlnIleTyThrProGlnGluArgHisSerIleArgValProGluSerGlyG 286
Db 134 CAATGANTTNCAGAACCAACCTCNGAGAAACCAANNAAGGTTNCGAATCGGAG 75
Qy 286 lueHisTyT 288
Db 74 GACAAATAT 67
RESULT 3
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTII
; TITLE OF INVENTION: INHIBITOR AND VASOPREPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
; FILE REFERENCE: Acty Docet No. 6586198 1242/48/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-593-5
Alignment Scores:
Pred. No.: 1.59e-29 Length: 3407
Score: 337.50 Matches: 99
Percent Similarity: 42.94% Conservative: 53
Best Local Similarity: 27.97% Mismatches: 127
Query Match: 20.09% Indels: 75
DB: 4 Gaps: 13
US-10-070-464-3 (1-310) x US-10-002-593-5 (1-3407)
Qy 5 lueAsnSerProLeuGlnHisLeuTyValValSerTyValAsnProGlyGlyVal 24
Db 1342 AAAGAAATGCCAGAGAGAGAAATCTTTATTAATAATCAACTATATGATACAAAGTG 1401
Qy 25 ThrArgLeuThrArgAcGlyTyThrSerHisSerCysIleSer---GlnHisCysAsp 43
Db 1402 ACATGCTC-----AGTTGTAGCTGAATCCGGAAGGTGTCTAG 1440

```

Oy 44 phepelleSerlySerAenglnlyAasnProHis-----Cys----- 57
Db 1441 TACTATTCTGTGTCATTCAGTAAGAGGAGAACTATTATCAGCTGAGTCCGGTCC 1500
Oy 58 ---ValSerLeuTyrIleuSerSerProGluAaspProThyrGlySerThrlYglu 76
Db 1501 GGTTCGCCCTCTATACCTTACACACCGCTGAATAAAGCGCTGAGCTCGTGA 1560
Oy 77 PheTrpAlaThrIleuAaspSerAlaGlyProLeuProAaspTyrThrProGlu--- 95
Db 1561 GAC---AATTCACTTGGATTAATAAG-----CTGCAGAAATCTCAGATGCCCTCCAA 1611
Oy 96 -----IlePheSerPheGluSerThrThrlGlyPheThrLeuTyrGlyMetLeuTyrLys 113
Db 1612 AAAGTGCATTCATTATTTTGAATGAACAAATTT---TGGTATCAGATGATCTTGCCCT 1668
Oy 114 ProHisAaspLeuGlnProGlyLysLysTyrProThrlValLeuPheIleTyrGlyPro 133
Db 1669 CCTCAT---TTGTGATTAATTCAGAAATATCTCTATCTATGATGTGTATGACAGGCCCA 1725
Oy 133 ----- 133
Db 1726 TGTAAGCAAAAAGACACACTGTCTTCAAGTGAAGTGGGCCACTTACCTTGCAAGACA 1785
Oy 133 ----- 133
Db 1786 GAAACATTATAGTAGTACTTGTATGACAGAGAGTGTATCCAGAGAGATAGATGC 1845
Oy 134 -----GlnGlyGlnIleGlnIleAaspGlnIleGlnGlyLeu 146
Db 1846 ATGCATGCATCAACAGAGAGACTGGAGACATTTGAAGATCAAAATTGAAGCAGCC 1905
Oy 147 GlnTyrLeuAlaSerArgTyrAaspPheIleAaspLeuAparValGlyIleHisGlyTyr 166
Db 1906 AGACAAATTT---TCAAAAATGGGATTTGTGGACCAACAAAGCAATTCGAATTTGGGGCTGG 1962
Oy 167 SerTyrGlyGlyTyrIleuSerLeuMetAlaLeuMetGlnArgSerAaspIlePheArgVal 186
Db 1963 TCATATGAGAGGTGACCTAATCAATGCTCTGGGATCGGAGAACTGGCGTTCAGATGT 2022
Oy 187 AlaIleAlaGlyAlaProValThrIleuTyrIlePheTyrAaspThrlYThrGlnuag 206
Db 2023 GGAATAGCGCTGCGCTGTATCCCGGTGGAGACTATGACTGACTGATACACAGAACT 2082
Oy 207 TyrMetGly-----HisProAaspGlnAenglnGlnGlyTyrTyrLeuGlySerValAla 224
Db 2083 TACATGGGTCTCCCAACTCCAGAGCAACCTTGACCATTAAGAAATTCACACAGTCATG 2142
Oy 225 MetGlnAlaGlyLysPheProSerGluProAaspArgLeuLeuHisGlyPheLeu 244
Db 2143 AGCAGAGCTGGAATAATTTT-----AAACAAGTTGAGTACCTCTTATTCATGGAACAGCA 2196
Oy 245 AaspGluAaspValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 264
Db 2197 GATGATTAACGTTTCACTTACAGCAGTCAAGATCTCCAAAGCCCTGGTCATGTGGA 2256
Oy 265 LysProTyrAaspLeuGlnIleTyrProGlnGlnuagHiserIleArgValProGluSer 284
Db 2257 GTGAGATTTCCAGCAATGTGGTATCTAGTGAAGACCAATGGAATGCTGACAGCAGCAGCA 2316
Oy 285 GlyGlnHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlu 298
Db 2317 CACCAACATATATATATACCAACATGAGCCACTCATTAACAA 2358

```

RESULT 4

PCT-US93-07923-1

Sequence 1, Application PC/TUS9307923

GENERAL INFORMATION:

APPLICANT: Morimoto, Chikao

APPLICANT: Schloesman, Stuart F.

APPLICANT: Tanaka, Toshikaji

TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

NUMBER OF SEQUENCES: 16

```

/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ COMPUTER: IBM PS/2 Model 502 or 555X
/ OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
/ SOFTWARE: WordPerfect (Version 5.0)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/07923
/ FILING DATE: 19930819
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/934,162
/ FILING DATE: 21-AUG-1992
/ APPLICATION NUMBER: 07/832,211
/ FILING DATE: 06-FEB-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Frazer, Janis K.
/ REGISTRATION NUMBER: 34,819
/ REFERENCE/DOCKET NUMBER: 00530/055002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 542-5070
/ TELEFAX: (617) 542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2924
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ PCT-US93-07923-1

```

Alignment Scores:

```

Pred. No.: 3,72e-29 Length: 2924
Score: 333.50 Matches: 99
Percent Similarity: 42.18% Conservative: 52
Best Local Similarity: 27.65% Mismatches: 124
Query Match: 19.85% Indels: 83
DB: 5 Gaps: 14

```

US-10-070-464-3 (1-310) x PCT-US93-07923-1 (1-2924)

```

Oy 5 LysAaspSerProLeuGlnHisIleuTyrValValSerTyrValAasnProGlyGluVal 24
Db 1277 AAAGAAATGCCAGAGAGAAAGAAATCTTATTAATAATC----- 1312
Oy 25 ThrArgLeuThrAaspArgGlyTyrIleSerHisSerCysHisIleSer----- 39
Db 1313 ---CAACTTAGTAC-----TATACAAAGTACATGCTTCAGTTGTGAGCTGAATCCG 1363
Oy 40 GlnHisCysAaspPheIleSerLysTyrSerAenglnlyAasnProHis----- 56
Db 1384 GAAAGGTCTACATCTATCTGTGTCATTCAGTAAGAGGCGCAAGTATTATACGCTGACA 1423
Oy 57 Cys-----ValSerLeuTyrLysLeuSerSerProGluAaspProThrCys 72
Db 1424 TGTTCGGGCTGCTGCTGCCCTCTATACCTTACACACAGCGTGAATGAAGGCGCTG 1483
Oy 73 LysThrLysGlnPheTrpAlaThrIleLeuAaspSerAlaGlyProLeuProAaspTyrThr 92
Db 1484 AGAGTCTCGAAGAC---AATTCACTTTGGAATAAATG-----CTGCAGAAATGCCAG 1534
Oy 93 ProProGlu-----IlePheSerPheGluSerThrThrlGlyPheThrLeuTyrGly 109
Db 1535 ATGCCCTCCAAAAAAGTGAATTCATTATTTTGAATGAACAAATTT---TGGTATCAG 1591
Oy 110 MetLeuTyrLysProHisAaspLeuGlnProGlyLysLysTyrProThrlValLeuPheIle 129

```

```

Db      1592 ATGATCTTGCCTCCATCAT---TTTGATTAATCCAGAAATATCCTTACTATTAGATGTG 1648
Qy      130 TyrGlyGlyPro-----
Db      1649 TATGACAGGCCCATGTAGTCAAAAAGACAGACTGTCTTCAAGCTGAATGGGCCACTTAC 1708
Qy      133 -----
Db      1709 CTTCGACAGCAGAAAAACATTATAGTAGTACCTTGAATGCGACAGAAAGTGTTACCA 1768
Qy      134 -----
Db      1769 GGAATATAGATCATGATGATGCAATCAACAGAACTGGAACTTTAAGTGAAGTCA 1828
Qy      143 ValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuSparValGly 162
Db      1829 ATTTGAAGCAGCCAGACAATTT---TCAAAATGGGATTTGTGACAAACAAAGCAATTGA 1885
Qy      163 ILeHsGlyTyrPheSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerA 182
Db      1886 ATTTGGGGCTGTCATATGAGAGGATCGTAACCTCATATGCTCTGGATCAGAAAGTGGC 1945
Qy      183 ILePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGly 202
Db      1946 GTGTTCAAGTGTGGAATTAAGCCGCGGCTGTATCCCGTGGAGTACTATGACTAGT 2005
Qy      203 TyrThrGluArgTyrMetGly-----HisProAspGlnAsnGluGlnGlyTyrTyrLeu 220
Db      2006 TACACAGAAAGTTACATATGGGTCTCCCACTCCAGAAAGAACCTTACACCTTACAAAT 2065
Qy      221 GlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeu 240
Db      2066 TCAACGTCATGAGCAGACCTGAAAATTTT---AACCAAGTTAGTACCTCTTAT 2119
Qy      241 HisGlyPheLeuAspGlnValHisPheAlaHisThrSerIleLeuSerPheLeu 260
Db      2120 CATGGAACAGCAGATATATACGTTCACTTACAGACAGCTCAGATCTCCAAAGCCCTG 2179
Qy      261 ValArgAlaGlyLeuProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280
Db      2180 GTGATGTTGGAGTGGATTTCACGAGCATGTGATATATGATGAACACATGGAATAGCT 2239
Qy      281 ValProGluSerGlyLysIleTyrGlyLeuHisLeuLeuHisTyrLeuGlnGlu 298
Db      2240 AGCAGACAGCAGACCAACATATATATACCCACATAGACCATTAACAA 2293

RESULT 5
US-09-462-284-1
; Sequence 1, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Michael
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Fungus
US-09-462-284-1

Alignment Scores:
Pred. No.: 5,79e-28 Length: 5496
Score: 327.00 Matches: 103

```

```

Percent Similarity: 41.45% Conservative: 40
Best Local Similarity: 29.86% Mismatches: 126
Query Match: 19.46% Indels: 76
DB: 4 Gaps: 11

US-10-070-464-3 (1-310) x US-09-462-284-1 (1-5496)

Qy      4 ThrLysAspSerProLeuGlnHisIleuTyrValValSerTyrValAsnProGlyGlu 23
Db      3146 ACTCAACACCAACAGCAGCCAGGCGCATCTCTACCTCCGCTCCAT---TCCAGTTTGGC 3202
Qy      24 ValThrArgLeuThrAspArgGlyTyrSer-----HisSerCysAlaSerGlnHis 41
Db      3203 GTACCCCGCTGTCGACACACCGTTGCGGCTACTGCTTCTTCTCCCGGAGAC 3262
Qy      42 CysAspPhePheIleSerTyrSerLeuAsnGlnLysAsnProHisCysValSerLeuTyr 61
Db      3263 TCGGGCTACTACATCTCTCATACGAGGCCACAGACCTTAC---CAGAACTCTAC 3319
Qy      62 LysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIle 81
Db      3320 ACGACCAACAGT-----ACCAACCACTC---CGCACATTC 3352
Qy      82 LeuAspSerAlaGlyProLeu-----ProAspTyrThrProGluIlePheSer 98
Db      3353 ACCGACAAAGCCAAAGTACTGAGCAATCAAGAGTATGATTCGCCCAACATCACCTTAC 3412
Qy      99 PheGlu-----SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAsp 116
Db      3413 TTGAGACTTCCCTCCCTCCGAGAAACCTTCAATGATGATGACAGGCTTACCCCGGG 3472
Qy      117 LeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLeuProGlnGlyGln 136
Db      3473 TTCTCCCGGATGAAGATGATCCCATTTTACCCCATACGCGGCCGACGAGCCCAA 3532
Qy      136 -----
Db      3533 GAAGTGACCAAGATGCGAAGCCCTGAATTCAAGCCTATGTGCTCCGACAGCGAA 3592
Qy      136 -----
Db      3593 CTGAGTACGTAACTGACCTGACATCCGACAAACCGCGACAGGTTTCAAGAGCCCAAGTTC 3652
Qy      137 -----
Db      3653 CCTCCGCGCGTACGCGCAACTCGGCGCTCCGGAAGCAGAAACAGATCTACGCGCG 3712
Qy      147 GlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyr 166
Db      3713 CAACAG--GCGGCAACATCCCTGGATGATGACAGACACATCGGACTTGGGGCTGG 3769
Qy      167 SerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgVal 186
Db      3770 AGTTTGGAGCTACTTGAACGAAAGCTCTCGAAGAGACAGCGCGCTTCAATTA 3829
Qy      187 AlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArg 206
Db      3830 GGAAGTCATACCGCCCTTCTGATCAGTGGCTTCTTCAAGTCAATGATGACGAGCGCC 3889
Qy      207 TyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlnSerValAlaMetGln 226
Db      3890 TACATGAAAGCCCTTCGACCAATGAGAGGCTCGAAGACCAAGCCGCTC---CGCAAG 3946
Qy      227 AlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGln 246
Db      3947 ACTGACGGGTTCAAGACGTGACGAGGCGGATCTTGATTCACACAGAAAGCGGCGACGAT 4006
Qy      247 AsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysPro 266
Db      4007 AACGTCATTTCCAGAACTCGGCTCGCTGGATCTCTCATGATGCGCATGCGC----- 4060
Qy      267 TyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGlu 286

```

Db 4061 -----GTCCTCTCGTAGAAGCTCCATTGCGCAATGTTCAAGACTGACAC 4108
Qy 287 H1sTYrGLuLeuH1s 291
Db 4109 GGAATCAGCTACCAT 4123
RESULT 6
US-08-230-491A-1
Sequence 1, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Galin-Chessa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELPE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, NO. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 Base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-230-491A-1
Alignment Scores:
Pred. No.: 1,866-25 Length: 2815
Score: 302.00 Matches: 93
Percent Similarity: 40.554 Conservative: 55
Best Local Similarity: 25.484 Mismatches: 139
Query Match: 17.984 Indels: 78
Gaps: 11
US-10-070-464-3 (1-310) x US-08-230-491A-1 (1-2815)
Qy 1 PheGLuGLyThrLysAspSerProLeuGLuH1sLeuTYrValValSerTYrValaAn 20
Db 1409 TTCAGAGTAACACAGATTCAGTCTTTATCTAGCAATGAATTTGAGAAATACCTGGA 1468
Qy 21 ProGLuGLuValThrArgLeuThrAspArgGLyTYrSerHisSer---CysCysAlIeSer 39
Db 1469 AGAAGAAACATCTACAGAAATTCAGATTCGAGAGCTATCTCCAAAGCAAGAGTGTCTTACT 1528
Qy 40 GlnH1s-----CysAspPhePheIleSerLysTYrSerAsnGlnLysaAn 54
Db 1529 TCCGATCTTAAGAAAGAAAGTCCAAATATTAACAAGCAAGTTTCAAGCACTAACGCCAAG 1588
Qy 55 ProH1sCysValSerLeuTYr-----LysLeuSerSerProGlnLysAspPro 70
Db 1589 TACTATGACCTTGTCTGCTAAGGCCAGGACATCCCATTTCCACCCCTTCATGATGAGAGC 1648

Qy 71 ThrCysLysThrIleGLuPheThrPalaThrIleLeuAspSerAlaGLyProLeuProAsp 90
Db 1649 ACTGATCAAGAA-----ATTAAATCTCTGAAAGAAACAAAGAAATTTGAAAT 1696
Qy 91 -----TYrThrProProGlnIlePheSerPheGLuSerThrIrnGLyPhe 105
Db 1697 GCTTGAATAATATCCAGCTGCTTAAGAGGAATTAAGAACTTGAAGATGAAAT 1756
Qy 106 ThrLeuTYrGLyMeLeuTYrLysProH1sAspLeuGlnProGLyLysLysTYrProH1r 125
Db 1757 ACTTTATGGTACAGAGTGAATCTCTCTCTCAATTTGACAGATCAAGAAAGATATCCCTTG 1816
Qy 126 ValLeuPheIleTYrGLyGLyPro----- 133
Db 1817 CTAAITCAAGTATATGATGATGCTCCGAGTCAAGATGTAAGTGTGTAATTCCTGTTAAT 1876
Qy 133 ----- 133
Db 1877 TGAGATCTTATCTTGAAGTAAGAAAGGATGTCATTCCTGCTGAGATGTCGAGGA 1936
Qy 134 -----GlnGLyGlnIleGlu 138
Db 1937 ACAGCTTTCGAAGTGACCAAACTCTATAGCAGTGTATGAAAGCTGGGTGTTATGAA 1996
Qy 139 IleAspAspGlnValGLuGLyLeuGlnTYrLeuAlaSerArgTYrAspPheIleAspLeu 158
Db 1997 GTTGAAGACCAAGATTAACAGCTGTCAGAAATTCATA--GAATGGCTTCATGTATGAA 2053
Qy 159 AspArgValGLyIleHisGLyTYrPseTYrGLyGLyTYrLeuSerLeuMetAlaLeuMet 178
Db 2054 AAAAGAAATGCATATGGGGCGTGCCTTATGAGATATGCTTCACTACATGCGCTTGCA 2113
Qy 179 GlnArgSerAspIlePheArgValAlaIleAlaGLyAlaProValThrLeuTrIlePhe 198
Db 2114 TCTGAACTGCTCTTTCAAAATGGGTATAGCAGTGGCTCCAGTCTCCAGTGGGAATAT 2173
Qy 199 TYrAspThrGLyTYrThGLuArgTYrMetGLyHisPro-----AspGlnAsnGLuGln 216
Db 2174 TACGGCTCTGTCTACACAGAGATTCATGGGCTCTCCCAAAAGAGATATATCTTGAG 2233
Qy 217 GLYTYrTYrLeuGLySerValAlaMetGlnAlaGLyLysPheProSerGLuProAsnArg 236
Db 2234 CACTATTAAGAAATTCACACTGTGATGCGCAAGAGAGAAATATTGCAAAATGATAGC----- 2287
Qy 237 LeuLeuLeuLeuH1sGLyPheLeuAspGLuAsnValHisPheAlaHisThrSerIleLeu 256
Db 2288 TATCTTTCATCCAGCGAAACAGCAGATGATATATGCACTTCAAAAACACAGCAGATT 2347
Qy 257 LeuSerPheLeuValArgAlaGLyLysProTYrAspLeuGlnIleTYrProGlnGLuArg 276
Db 2348 GCTAAAGCTCTGTATTAATGACAAAGTGAATTTCCAGGCAATGTGTACTTGACCAAGAC 2407
Qy 277 HisSerIleArgValProGLuSerGLy-----GlnH1sTYrGLuLeuH1sLeu 293
Db 2408 CACGGCTTA-----TCCGGCTGTCCAGAAACCACTTATTAACCCACATGACC 2455
Qy 294 HisTYrLeuGlnGlu 298
Db 2456 CACTTCTTAAGCAG 2470
RESULT 7
US-08-619-280A-1
Sequence 1, Application US/08619280A
Patent No. 5767242
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felpe & Lynch

STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 Base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-619-280A-1
Alignment Scores:
Pred. No.: 1,86e-25 Length: 2815
Score: 302.00 Matches: 93
Percent Similarity: 40.55% Conservative: 55
Best Local Similarity: 25.48% Mismatches: 139
Query Match: 17,98% Indels: 78
DB: 1 Gaps: 11
US-10-070-464-3 (1-310) x US-08-619-280A-1 (1-2815)
QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleuLeuTyValValSerTyValaAn 20
DB 1409 TTCAGAGTACACAGAGATTCACCTTTTATTCAGCAATGAATTTGAAGATACCCCTGGA 1468
QY 21 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSer---CysCysIleSer 39
DB 1469 AGAAGAAACATCTACAGAAATTCAGATTCGAAAGCTATCTCCAGCAAGAAAGTGTGTACT 1528
QY 40 GlnHis-----CysAspPhePheIleSerLysTyrSerGlnLysaAn 54
DB 1529 TGGCACTTAAGAAAGAGTCCCAATATTACACAGCAAGTTTCAGCCGCTACGCCAAG 1588
QY 55 ProHisCysValSerLeuTyr-----LysLeuSerSerProGluAspArgPro 70
DB 1589 TACTATGACACTTCTGCTACGCGCCAGGATCCCATTCACCCCTTATATGAGACGC 1648
QY 71 ThrCysLysThrLysGluPheThrAlaThrIleLeuAspSerLysGlyProLeuProAsp 90
DB 1649 ACTGATCAAGAA-----ATTAAATCTCTGAGAAAGAAACAAAGAAATTTGAAAAAT 1696
QY 91 -----TyrThrProProGluIlePheSerPheGluSerThrArgLysPhe 105
DB 1697 GCTTTAAAAATATTCAGCTGCTTAAAGAGAAATTAAGAAATTAAGTAAATTAAT 1756
QY 106 ThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThr 125
DB 1757 ACTTATGAGTACAGAAAGATGATTTCTCTCTCAATTTGACAGAGATCAAGAAAGATCCCTTG 1816
QY 126 ValLeuPheIleTyrGlyLysPro----- 133
DB 1817 CTAAATCAAGTATATGATGTGTCTCTGCAAGTCAAGTGAAGTGAAGTGTATTTGCTGTAAAT 1876

QY 133 ----- 133
DB 1877 TGGATATCTTATCTTCAAGTAAGAGGATGTCATTCCTGTGAGTGCAGGA 1936
QY 134 -----GlnGlyGlnIleGlu 138
DB 1937 ACAGCTTTCCAAAGTACAAATCTCTATGACGTATGCAAGCTGGGTGTTATGAA 1996
QY 139 IleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeu 158
DB 1997 GTTGAAGACCAAGTTTACAGTGTACAGAAATTCATA---GAATGGGTTTCATTGATGA 2053
QY 159 AspArgValGlyIleHisGlyTyrPheSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMet 178
DB 2054 AAAAGATAGCCATATGAGGAGTGTCTCTATGAGGATAGAGTTTCATCAGCTGGCCTTGCA 2113
QY 179 GlnArgSerAspPheArgValAlaIleAlaGlyAlaProValThrLeuTyrIlePhe 198
DB 2114 TCTGAACTGTCTTTTCAATGTGTATAGACGTGCTCCAGCTCCAGCTGGAATAT 2173
QY 199 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro-----AspGlnAsnGln 216
DB 2176 TACGGTCTGTGTACACAGAGATTCATGGGTCTCCAAAGATGATATCTTGAG 2233
QY 217 GlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArg 236
DB 2234 CACTATAGAAATTCACATGTCGTGAGCAGAGAGAAATATTTCAGAAATGTAGAC----- 2287
QY 237 LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 256
DB 2288 TATCTTCATATCCAGGAAACAGCAGATGATTAATGTCACTTTCAAACCTCAGCAGAT 2347
QY 257 LeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGlnArg 276
DB 2348 GCTAAAGCTGTGTATATGACAAAGATGATTCACAGCAATGCGTACCTGACCAAGAC 2407
QY 277 HisSerIleArgValProGluSerGly-----GlnHisTyrGluLeuHisLeuLeu 293
DB 2408 CACGGCTTA-----TCCGGCTGTGTCCAGCAACCACTTATATACCCACATGACC 2455
QY 294 HisTyrLeuGlnGlu 298
DB 2456 CACTTCTTAAGCAG 2470
RESULT 8
US-08-940-391-1
Sequence 1, Application US/08940391
Patent No. 5965373
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
TITLE OF INVENTION: Ret(19, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280

```

/ FILING DATE: 18-MARCH-1996
/ APPLICATION NUMBER: 08/230,491
/ FILING DATE: 20-APRIL-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hanson, No. 5965373man D.
/ REGISTRATION NUMBER: 30,946
/ REFERENCE/DOCKET NUMBER: LUD 5330.1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 688-9200
/ TELEFAX: (212) 638-3884
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2815 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/
US-08-940-391-1

Alignment Scores:
Pred. No.: 1,866-25 Length: 2815
Score: 302.00 Matches: 93
Percent Similarity: 40.554 Conservative: 55
Best Local Similarity: 25.484 Mismatches: 139
Query Match: 17.984 Indels: 78
DB: Gaps: 11

US-10-070-464-3 (1-310) x US-08-940-391-1 (1-2815)

OY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuValValSerThrValaen 20
DB 1409 TTCAGAGTACACAGGATTCAGTTTATTCTGACGATGAATTTGAGAAATACCTGGA 1468
OY 21 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSer---CysCysHisSer 39
DB 1469 AGAAGAACATCTACGATTCAGATTGACATGGAAGCTATCCCAAGCAAGAGTGTACT 1528
OY 40 GluHis-----CysAspPhePheHisSerLysTyrSerAsnGlnLysAsn 54
DB 1529 TGCCATCTTAAAGAAAGAGGCGCAATATTACACAGCAAGTTTCAGCAATACGCAAG 1588
OY 55 ProHisCysValSerLeuTyr-----LysLeuSerSerProGluAspAspPro 70
DB 1589 TACTATGCACTGTCTGCTACGAGCCAGCATCCCATTTCCACCTTCATATGAGACGC 1648
OY 71 ThrCysLysThrLysGluPheThrPalaThrLysLeuAspSerAlaGlyProLeuProAsp 90
DB 1649 ACTGATCAAGAA-----ATTAAATCCCTGGAAGAAAGAAAGCAAGATTTGAAAT 1696
OY 91 -----TyrThrProProGluLysPheSerPheGluSerThrGlyPhe 105
DB 1697 GCTTGAATAATATCCAGCTGCTAAAGAGAAATTAAGAACTTGAAATGATGAAT 1756
OY 106 ThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysThrProThr 125
DB 1757 ACTTATAGTACAAAGATGATTTCTTCTCTCAATTGACATTCAGATCAAGAGATCCCTTG 1816
OY 126 ValLeuPheLeuTyrGlyGlyPro----- 133
DB 1817 CTAATTCAAGGTATAGTGTCTCCGACATGACAGTGAAGTCTGTAATTGCTGTAAT 1876
OY 133 ----- 133
DB 1877 TGGATATCTTATCTTGAAGTAAGAGGATGTCATTCCTTGATGAGATGTCAGAGA 1936
OY 134 -----GlnGlyGlnLeuGlu 138
DB 1937 ACAGCTTTCCAAGGTGACAAACTCTCTATGACAGTGTATCGAAAGCTGGGTTTATATA 1996
OY 139 IleAspArgGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeu 158
DB 1997 GTTGAAGACCAAGATTAAGCTGTACAGAAATTCATA---GAATGGGTTTCAATTGATA 2053
OY 159 AsparGValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMet 178

```

```

DB 2054 AAAAGATAGCCATATGGGGCTGCTCTATGAGAGATACGTTTCATCACTGGCCCTTGA 2113
OY 179 GlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuPheIlePhe 198
DB 2114 TCTGAAGCTGCTTTTCAAAGTGTATAGACAGTGGCTCCAGCTCCAGCTGGGAATAT 2173
OY 199 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro-----AspGlnAsnGluGln 216
DB 2174 TACGGCTGTCTTACACAGAGATTCATGAGTCTCCCAAGCAAGATATATATCTTGAG 2233
OY 217 GlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArg 236
DB 2234 CACTATTAAGAAATTCAGCTGTATGCGACAGACAGCAAAATATTTCAGAAATATAGAC----- 2287
OY 237 LeuLeuLeuLeuHisGlyPheLeuAspGlnValHisPheAlaHisThrSerIleLeu 256
DB 2288 TATCTTCATCCACGGAACAGCAGATGATTAATGTGCATTTCAAACACAGCAGAT 2347
OY 257 LeuSerPheLeuValAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArg 276
DB 2348 GCTAAAGCTCTGTTAATGACAAAGTGAATTTCCAGGCAATGGTACTCTGACCAAGAC 2407
OY 277 HisSerIleArgValProGluSerGly-----GluHisTyrGluLeuHisLeu 293
DB 2408 CACGCTTA-----TCCGAGCTGTCCAGAACCACTTATACACCCACATGACC 2455
OY 294 HisTyrLeuGlnGlu 298
DB 2456 CACTTCTTAAAGCAG 2470

RESULT 9
US-08-699-103B-1/C
/ Sequence 1, Application US/08699103B
/ Patent No. 6107462
/ GENERAL INFORMATION:
/ APPLICANT: Hampe, Jasper D.
/ TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
/ TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Fish & Richardson P.C.
/ STREET: 2200 Sand Hill Road, Suite 100
/ CITY: Menlo Park
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94025
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FaasSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/699,103B
/ FILING DATE: 16-AUG-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/002,581
/ FILING DATE: 17-AUG-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Green, Grant D.
/ REGISTRATION NUMBER: 31,259
/ REFERENCE/DOCKET NUMBER: 09272/005001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650/322-5070
/ TELEFAX: 650/854-0875
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4982 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA

```

US-08-699-103B-1

Alignment Scores:

Pred. No.:	6,19e-21	Length:	4982
Score:	267.00	Matches:	67
Percent Similarity:	42.65%	Conservative:	23
Best Local Similarity:	31.75%	Mismatches:	61
Query Match:	15.89%	Indels:	60
DB:	3	Gaps:	6

US-10-070-464-3 (1-310) x US-08-699-103B-1 (1-4982)

```

OY      123  TyrProthValLeuPheIleTyrglyProgin-----
DB      4941  TATCCTGATTTTCTTTCATATGCGGAGCGAATTCACAAGTTGTCAAAACGTTT 4882
OY      134
DB      4881  TCCGTAGATTAAATGAAGTGTAGCTTCAATTAACGAATTGTTGTTGAC 4822
OY      135
DB      4821  GGTGCTGACTGCTTCAAGGTCAAGACTTAGATCCCTTGTTCGATAGCTCGGT 4762
OY      136  GlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPhe 155
DB      4761  GATTACGAGCGCCGCGACCAATATCTCGCGCTTCTTATATGTTCT--TTAAGTTT 4705
OY      156  IleAspLeuAspArgValGlyIleHisGlyTyrSerTyrglyTyLeuSerLeuMet 175
DB      4704  GTTGATCCGCAAAAGATTTCCATTATTTGTTGTCATACGGGGGTACTGACACTAAAA 4645
OY      176  AlaLeuMetGlnArgSerAsp--IlePheArgValAlaIleAlaGlyAlaProValThr 194
DB      4644  ACTTTGAGAAAGATGCGCGAAGACATTTCAATACGGGATGTCAGTTCCGCGAGTAACC 4585
OY      195  LeuTyrIlePheTyAspThrGlyTyrThrGluArgTyMetGlyHisProAspGlnAsn 214
DB      4584  GACTGAGACTTTTACATTCATCTTTTACTGAGAGGATACATACCTCTCAAGAAAA 4525
OY      215  GluGlnGlyTyrTyLeuGlySerVal-----AlaMetGlnAlaGluLysPhe 230
DB      4524  TTTGATGATACGTATGATCAACGCTTCAATATGTCAGCTTTG----- 4480
OY      231  ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 250
DB      4479  ---GCAACAAGCAAAATGATTTTGTGATGACGACGGAACAGAGATGATTAACCTTCACTT 4423
OY      251  AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly---LysProTyAspLeu 269
DB      4422  CAAATTCCTTAAGATTTCTGACCTTTTGATCTTAATGATGTGTGAAAAATTAATGACGTC 4363
OY      270  GlnIleTyProGlnGluArgHisSerIleArg 280
DB      4362  CACGTTTCTCTGACCTCAGATCATAGTATTAAGA 4330

```

RESULT 10

US-09-229-059-1/c

Sequence 1, Application US/09229059

Patent No. 6333172

GENERAL INFORMATION:

APPLICANT: Rine, Jasper D.

APPLICANT: Hampton, Randolph

TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING

TITLE OF INVENTION: CHOLESTEROL SYNTHESIS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 2200 Sand Hill Road, Suite 100

CITY: Menlo Park

STATE: CA

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/229,059

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/599,103

FILING DATE: 16-AUG-1996

APPLICATION NUMBER: 60/002,581

FILING DATE: 17-AUG-1995

ATTORNEY/AGENT INFORMATION:

NAME: Green, Grant D.

REGISTRATION NUMBER: 31,259

REFERENCE/DOCKET NUMBER: 09272/005001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/322-5070

TELEFAX: 650/854-0875

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4982 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-229-059-1

Alignment Scores:

Pred. No.:	6,19e-21	Length:	4982
Score:	267.00	Matches:	67
Percent Similarity:	42.65%	Conservative:	23
Best Local Similarity:	31.75%	Mismatches:	61
Query Match:	15.89%	Indels:	60
DB:	4	Gaps:	6

US-10-070-464-3 (1-310) x US-09-229-059-1 (1-4982)

```

OY      123  TyrProthValLeuPheIleTyrglyProgin-----
DB      4941  TATCCTGATTTTCTTTCATATGCGGAGCGAATTCACAAGTTGTCAAAACGTTT 4882
OY      134
DB      4881  TCCGTAGATTAAATGAAGTGTAGCTTCAATTAACGAATTGTTGTTGAC 4822
OY      135
DB      4821  GGTGCTGACTGCTTCAAGGTCAAGACTTAGATCCCTTGTTCGATAGCTCGGT 4762
OY      136  GlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPhe 155
DB      4761  GATTACGAGCGCCGCGACCAATATCTCGCGCTTCTTATATGTTCT--TTAAGTTT 4705
OY      156  IleAspLeuAspArgValGlyIleHisGlyTyrSerTyrglyTyLeuSerLeuMet 175
DB      4704  GTTGATCCGCAAAAGATTTCCATTATTTGTTGTCATACGGGGGTACTGACACTAAAA 4645
OY      176  AlaLeuMetGlnArgSerAsp--IlePheArgValAlaIleAlaGlyAlaProValThr 194
DB      4644  ACTTTGAGAAAGATGCGCGAAGACATTTCAATACGGGATGTCAGTTCCGCGAGTAACC 4585
OY      195  LeuTyrIlePheTyAspThrGlyTyrThrGluArgTyMetGlyHisProAspGlnAsn 214
DB      4584  GACTGAGACTTTTACATTCATCTTTTACTGAGAGGATACATACCTCTCAAGAAAA 4525
OY      215  GluGlnGlyTyrTyLeuGlySerVal-----AlaMetGlnAlaGluLysPhe 230
DB      4524  TTTGATGATACGTATGATCAACGCTTCAATATGTCAGCTTTG----- 4480
OY      231  ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 250
DB      4479  ---GCAACAAGCAAAATGATTTTGTGATGACGACGGAACAGAGATGATTAACCTTCACTT 4423
OY      251  AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly---LysProTyAspLeu 269
DB      4422  CAAATTCCTTAAGATTTCTGACCTTTTGATCTTAATGATGTGTGAAAAATTAATGACGTC 4363
OY      270  GlnIleTyProGlnGluArgHisSerIleArg 280
DB      4362  CACGTTTCTCTGACCTCAGATCATAGTATTAAGA 4330

```


Db 4479 ---GCACAAAGCAATAGATTGTTGTGAGCAGGAAAGAGATGATAAGTTCACTTT 4423
Qy 251 AlaHisThrSerIleuLeuSerPheLeuValArgAlaGly---LysProTyrAspLeu 269
Db 4422 CAAATTCCTTAAGATTCTCGACCTTTGGATCTAAATGCTGTGAAAAATTATGACGTC 4363
Qy 270 GlnIleTyrProGlnGluArgHisSerIleArg 280
Db 4362 CACGCTCTTCTGACTCAGATCATAGTATAGGA 4330
RESULT 11
US-09-628-133-1/c
Sequence 1, Application US/09628133
Patent No. 6531292
GENERAL INFORMATION:
APPLICANT: Rine, Jasper D.
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/628,133
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,103
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 09272/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/854-0875
TELEFAX: 650/322-5070
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4982 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-628-133-1
Alignment Scores:
Pred. No.: 6.19e-21 Length: 4982
Score: 267.00 Matches: 67
Percent Similarity: 42.65% Conservative: 23
Best Local Similarity: 31.75% Mismatches: 61
Query Match: 15.89% Indels: 60
DB: 4 Gaps: 6
US-10-070-464-3 (1-310) x US-09-628-133-1 (1-4982)
Qy 123 TyrProThValLeuPheIleTyrGlyGlyProGln----- 134
Db 4941 TATCCTGATTTTCTTTCGATATGGGAGCCGAATTTCACACAGTTGCAAAAGCTTT 4882
Qy 134 ----- 134
Db 4881 TCCGATGATTTAATGAGGTGAGCTTCACATTAAAGCAATTGTAGTTGTTGAC 4822
Qy 135 -----Gly 135

Db 4821 GGTGCTGATACCTGCTTCAAGGTCAAGACTTTAGATCCCTTGTCCGATAGGCTCGGT 4762
Qy 136 GlnIleGluIleAspArgIleValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPhe 155
Db 4761 GATTACGAGGCCCGCAGCAAAATATGTCGGCTTCTTTATAGCTTCT---TTAACTTTT 4705
Qy 156 IleAspLeuAspArgValGlyIleHisGlyTyrPheTyrGlyTyrLeuSerLeuMet 175
Db 4704 GTTGATCCCGAATAAGATTTCTTATTTGGTGTCTACCGGGGCTTACCTGACACTATAA 4645
Qy 176 AlaLeuMetGlnArgSerAsp---IlePheArgValAlaIleAlaGlyAlaProValThr 194
Db 4644 ACTTGAGAAAGATGCGGAGACACATTCAATATGCGATGACAGTTGGCGCAGTAC 4585
Qy 195 LeuTyrIlePheTyrAspThrArgTyrTyrGlnValArgTyrMetGlyHisProAspGlnAsn 214
Db 4584 GACTGAGATTTTATCATTCTGTGTTTACTGAGAGGTATCATGATCTCTCAAGAAAC 4525
Qy 215 GluGlnGlyTyrTyrLeuGlySerVal-----AlaMetGlnAlaGluTyrPhe 230
Db 4524 TTTGATGATACGTAGATCAAGCCGTTTCATATGTCAGTCTTGTG----- 4480
Qy 231 ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluLeuValHisPhe 250
Db 4479 ---GCACAAAGCAATAGATTGTTGTGAGCAGGAAAGAGATGATAAGTTCACTTT 4423
Qy 251 AlaHisThrSerIleuLeuSerPheLeuValArgAlaGly---LysProTyrAspLeu 269
Db 4422 CAAATTCCTTAAGATTCTCGACCTTTGGATCTAAATGCTGTGAAAAATTATGACGTC 4363
Qy 270 GlnIleTyrProGlnGluArgHisSerIleArg 280
Db 4362 CACGCTCTTCTGACTCAGATCATAGTATAGGA 4330
RESULT 12
US-09-221-017B-253
Sequence 253, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H

```
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 253:
SEQUENCE CHARACTERISTICS:
LENGTH: 543 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...543
US-09-221-017B-253

Alignment Scores:
Pred. No.: 6.01e-22 Length: 543
Score: 263.00 Matches: 51
Percent Similarity: 58.99% Conservative: 31
Best Local Similarity: 36.69% Mismatches: 53
Query Match: 15.65% Indels: 4
DB: Gaps: 2

US-10-070-464-3 (1-310) x US-09-221-017B-253 (1-543)
QY 135 GYGLHLLGLLLEGLLEASPAGLNLGLGLYLEGLNLYRLEUHLASERARGTYRASP 154
    |||||
DB 136 GGGCAGCCGAGATGGCCGATCATGTGCGGTGATTTCTCAAGACCAA--TCA 192
QY 155 PHEILASPLEUAPRAGVALIGYLIHISGLYTRPSERYRGLYTYRLEUSETLEU 174
    |||||
DB 193 TGGTGATGTCCTGATGAATAGAGTAGACGCTGAGCTATGCTTATGACTACG 252
QY 175 METALAEUWETGLNAGSERASPILEPHEARGVALALALEAGLAPROVALTHR 194
    |||||
DB 253 AATCTGATGCTTACGACGCGCATGCTTCAAGTCGAGTAGCCGCGGCTGTCTATA 312
QY 195 LEUTYRILEPHEYRAPHRYRGLYTYRTHGLUARGTYRMEGLYHISPROASPLASN 214
    |||||
DB 313 GACTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
QY 215 GLUINGLYTYRILEUGLYSERVALALAMERGLNLAULUYSRHEPROSERGIUPRO 234
    |||||
DB 373 CCCGAAGATACGATGCTGCCAACCTGCTCAAGACGCGGTGATCTG-----AAA 423
QY 235 ASNARGLEULEUENLEUHLISGLYPHELEUASPLUASNVALHISPLEALHISTHRSER 254
    |||||
DB 424 GGAAGACTTATGCTGATTCATGAGAGCATGATCCGCTGATGATGAGCAGCATTCCTC 483
QY 255 ILEULEUSERPHELEUVALARGALGLYSEPROTYRASPLEUGLNIETYPRO 273
    |||||
DB 484 CTTTCTTGATGCTTCCGTGAAGCAGCACCACTATCTGACTTACGTCTATCCG 540

RESULT 13
US-09-221-017B-646/C
Sequence 646, Application US/09221017B
GENERAL INFORMATION:
APPLICANT: ROSS, BRUCE C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA
```

```
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 646:
SEQUENCE CHARACTERISTICS:
LENGTH: 657 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...657
US-09-221-017B-646

Alignment Scores:
Pred. No.: 1.61e-20 Length: 657
Score: 252.00 Matches: 56
Percent Similarity: 42.78% Conservative: 24
Best Local Similarity: 29.95% Mismatches: 57
Query Match: 15.00% Indels: 50
DB: Gaps: 3

US-10-070-464-3 (1-310) x US-09-221-017B-646 (1-657)
QY 91 TYRTHPROGSLULEPHESERPHEGLUSERTHRTHGLYPHERHLEUHYRGLYMET 110
    |||||
DB 566 TACAACCCGAAGAGATTACCATATCAAACTCAATCCGGCTTGAACATGATGCTGG 507
QY 111 LEUTYRILEPHEPROHISAPLEUGLNPFGLYLYSELYRTPROTHRVALLLEUPHE 130
    |||||
DB 506 ATCGTAGAGCTTATGATTCGATCCCTCGCCACTATCTGCTCGATGATGATGATGAT 447
QY 131 GYGLYPLROGLNLYGLN----- 136
    |||||
DB 446 AGCGGTCCCACTCCAGCAGGATTGATGCTATTCATTGCGATGGGAACACTACTT 387
QY 136 ----- 136
DB 386 GCATGAAGATTACGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 327
```

```

Qy 137 -----11eglu1leaspaapglu144
Db 326 GAATGCCGCAAGTGTACTACATGCAACTCGGTATTCGAAAGGATGATCAGATACGA 267
Qy 145 G1yleug1n1yrlu1leaspaapglu144
Db 266 GCGGCGACTGCTGATG---GACACGCTGCCCTATGATGATGATGATGATGATGATG 210
Qy 165 G1ytrpser1yrlu1yrlu1leaspaapglu144
Db 209 GGGTGGACTGATGCGGCTATACACACTATGATGATGATGATGATGATGATGATGATG 150
Qy 185 ArgVal1a1leal1a1yrlu1leaspaapglu144
Db 149 AAAGCGGGGATGCGGCTGCTGCTGCGAGACTGCGGCTTCTACATGATGATGATGATG 90
Qy 205 G1u1ytrpser1yrlu1yrlu1leaspaapglu144
Db 89 GAACCTTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 30
Qy 223 ---Val1a1leal1a1yrlu1leaspaapglu144
Db 29 GATGCGCAAGCAATTAACA 9

RESULT 14
US-09-389-681-428
; Sequence 428, Application US/09389681A
; Patent No. 6518237
; GENERAL INFORMATION:
; APPLICANT: Yuqin, Jiang
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C3
; CURRENT FILING DATE: 1999-09-02
; CURRENT APPLICATION NUMBER: US/09/389,681A
; NUMBER OF SEQ ID NOS: 463
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 428
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-389-681-428

Alignment Scores:
Pred. No.: 1.09e-15 Length: 535
Score: 210.00 Matches: 50
Percent Similarity: 42.77% Conservative: 21
Best Local Similarity: 30.12% Mismatches: 45
Query Match: 12.50% Indels: 50
DB: 4 Gaps: 3

US-10-070-464-3 (1-310) x US-09-389-681-428 (1-535)
Qy 114 ProH1a1leal1a1yrlu1leaspaapglu144
Db 15 CCTCTCAATTTGACAGATCAAGAGATGATGATGATGATGATGATGATGATGATG 74
Qy 133 -----
Db 75 TGCACTCAGAGTGAAGTGTGATTTGCTGTAATGATGATGATGATGATGATGATG 134
Qy 133 -----
Db 135 GAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 194
Qy 134 -----
Db 195 CTCTATGAGTGTATGAAAGCTGCTGTTATGAGTGAAGTGAAGTGAAGTGAAGTGA 254
Qy 147 G1n1yrlu1leaspaapglu144
Db 147 G1n1yrlu1leaspaapglu144

```

```

Db 255 AGAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 311
Qy 167 Ser1yrlu1yrlu1leaspaapglu144
Db 312 TCTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 371
Qy 187 Ala1leal1a1yrlu1leaspaapglu144
Db 382 GGTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 431
Qy 207 Tyrlu1yrlu1leaspaapglu144
Db 432 TCTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 491
Qy 225 MetG1n1yrlu1leaspaapglu144
Db 492 GCAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 509

RESULT 15
US-09-620-405B-428
; Sequence 428, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqin
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C8
; CURRENT APPLICATION NUMBER: US/09/620,405B
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 495
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 428
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-620-405B-428

Alignment Scores:
Pred. No.: 1.09e-15 Length: 535
Score: 210.00 Matches: 50
Percent Similarity: 42.77% Conservative: 21
Best Local Similarity: 30.12% Mismatches: 45
Query Match: 12.50% Indels: 50
DB: 4 Gaps: 3

US-10-070-464-3 (1-310) x US-09-620-405B-428 (1-535)
Qy 114 ProH1a1leal1a1yrlu1leaspaapglu144
Db 15 CCTCTCAATTTGACAGATCAAGAGATGATGATGATGATGATGATGATGATGATG 74
Qy 133 -----
Db 75 TGCACTCAGAGTGAAGTGTGATTTGCTGTAATGATGATGATGATGATGATGATG 134
Qy 133 -----
Db 135 GAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 194
Qy 134 -----
Db 195 CTCTATGAGTGTATGAAAGCTGCTGTTATGAGTGAAGTGAAGTGAAGTGAAGTGA 254
Qy 147 G1n1yrlu1leaspaapglu144
Db 255 AGAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 311

```

```
QY 167 SerTYGlyGlyTYrleuSerleuMetAlaleuMetGlnArgSerAspIlePheArgVal 186
Db 312 TCCTATGAGAGATACGTTTCATCATCGGCCCTTGCACTGGAACCTGCTTTCAAAATGT 371
QY 187 AlaIleAlaGlyAlaProValThrleuThrIlePheTYrAspThrGlyTYrThrGluArg 206
Db 372 GGTATATGACGTGGCTTCAGTCTCCAGCTGGAAATATTACCGCTGTCTACACAGAGAGA 431
QY 207 TYrMetGlyHisPro-----AspGlnAsnGluGlnGlyTYrTYrleuGlySerValAla 224
Db 432 TTCATGGGCTCCCAACAAGGATGATTAATCTTGAGCACTATAGAAATTCACACTGTGATG 491
QY 225 MetGlnAlaGluLysPhe 230
Db 492 GCAAGAGCAGAAATATTTC 509
```

Search completed: October 16, 2003, 03:30:32
Job time : 70.5672 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: October 16, 2003, 00:05:32 ; Search time 231.77 Seconds
(without alignments)
3512.533 Million cell updates/sec

Title: US-10-070-464-3
Perfect score: 1680
Sequence: 1 FEGTQDPLEHLLVYVSYVN.....HLHLYLQNTGSRIALKVI 310

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1750203 seqs, 1313063994 residues
Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-O=/cgn2_1/USPTO/spool/US10070464/runat_15102003_113555_24924/app_query.fasta_1.2652
-DB=Published_Applications_NA -QMT=fastcap -SUFFIX=rmpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pcct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10070464 @CGN 1.1 845 @runat_15102003_113555_24924
-NCPU=6 -ICPU=3 -NO MMAR -LARGEQUERY -NEG SCORES=0 -WAIT -DSRPOOL=100 -XGAPEXT=0.5
-LONGLOG -DEVTIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1645.5	97.9	2649	US-10-054-776-1	Sequence 1, Appl1

2	1645.5	97.9	2649	12	US-10-170-789-39	Sequence 39, Appl1
3	1645.5	97.9	2671	10	US-09-976-674-2	Sequence 2, Appl1
4	1645.5	97.9	3143	12	US-10-170-789-37	Sequence 37, Appl1
5	1644.5	97.9	4576	10	US-09-976-674-20	Sequence 20, Appl1
6	1625.5	96.8	4509	10	US-09-976-674-14	Sequence 14, Appl1
7	1625.5	96.8	4829	10	US-09-976-674-12	Sequence 12, Appl1
8	1358.5	80.9	4685	10	US-09-976-674-22	Sequence 22, Appl1
9	1352	80.5	4523	10	US-09-976-674-8	Sequence 8, Appl1
10	1093.5	65.1	2617	10	US-09-976-674-4	Sequence 4, Appl1
11	1093.5	65.1	4219	10	US-09-976-674-28	Sequence 28, Appl1
12	1093.5	65.1	4302	10	US-09-976-674-24	Sequence 24, Appl1
13	1086.5	64.7	4180	10	US-09-976-674-36	Sequence 36, Appl1
14	1086.5	64.7	4263	10	US-09-976-674-34	Sequence 34, Appl1
15	1032.5	61.5	2801	13	US-10-098-841-100	Sequence 100, App
16	1032.5	61.5	3262	13	US-10-098-841-82	Sequence 83, Appl1
17	872.5	51.9	4076	10	US-09-976-674-32	Sequence 32, Appl1
18	872.5	51.9	4159	10	US-09-976-674-30	Sequence 30, Appl1
19	865.5	51.5	4037	10	US-09-976-674-40	Sequence 40, Appl1
20	865.5	51.5	4120	10	US-09-976-674-38	Sequence 38, Appl1
21	662.5	39.4	561	11	US-09-764-891-877	Sequence 877, App
22	545	32.4	502	11	US-09-918-985-19585	Sequence 19585, A
23	344.5	20.5	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
24	344.5	20.5	4835	14	US-10-165-603-5	Sequence 5, Appl1
25	337.5	20.1	3407	12	US-10-423-714-5	Sequence 5, Appl1
26	337.5	20.1	3407	13	US-10-002-593-5	Sequence 5, Appl1
27	337.5	20.1	3407	14	US-10-155-603-6	Sequence 6, Appl1
28	310	18.5	2130	14	US-10-156-761-3131	Sequence 3131, Ap
29	310	18.5	9025608	14	US-10-156-761-1	Sequence 1, Appl1
30	302	18.0	2366	12	US-10-101-510-683	Sequence 683, Appl1
31	302	18.0	2788	12	US-10-269-909-22	Sequence 22, Appl1
32	302	18.0	2814	10	US-09-962-832-108	Sequence 108, App
33	302	18.0	2814	10	US-09-954-456-1148	Sequence 1148, Ap
34	302	18.0	2814	10	US-09-954-571-367	Sequence 367, App
35	302	18.0	2814	12	US-10-101-510-30	Sequence 30, Appl1
36	302	18.0	2814	12	US-10-301-822-54	Sequence 54, Appl1
37	302	18.0	2814	12	US-09-873-367C-499	Sequence 499, App
38	302	18.0	2814	14	US-10-177-253-135	Sequence 135, App
39	302	18.0	2815	9	US-09-265-606-1	Sequence 1, Appl1
40	302	18.0	3138	14	US-10-198-866-13171	Sequence 13171, A
41	302	18.0	3224	12	US-10-240-965-117	Sequence 117, App
42	289	17.2	2812	14	US-10-084-817-24	Sequence 24, Appl1
43	289	17.2	4797	11	US-09-764-891-7074	Sequence 7074, Ap
44	277	16.5	2388	12	US-09-870-133-3	Sequence 3, Appl1
45	277	16.5	2388	14	US-10-160-501-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-10-054-776-1
Sequence 1, Application US/10054776
Publication No. US20030165818A1
GENERAL INFORMATION:
APPLICANT: Mark Robert Edbrooke
APPLICANT: Alan Peter Lewis
TITLE OF INVENTION: NOVEL PROTEIN
FILE REFERENCE: OG1042US
CURRENT APPLICATION NUMBER: US/10/054,776
CURRENT FILING DATE: 2002-01-23
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2649
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1) .. (2649)
US-10-054-776-1
Alignment Scores: 1.47e-209 Length: 2649
Pred. No.: 1645.50 Matches: 310

Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
DB: 12 Gaps: 1

US-10-070-464-3 (1-310) x US-10-054-776-1 (1-2649)

```
QY 1 PheGluGlyThrLysAspSerProLeuGluHisLysLeuValValSerTyValaen 20
DB 1570 TTTGAAGGACCAAGACTCCCTTTAGAGCATCACTTAGTACGTAGTACGTAAT 1629
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
DB 1630 CCTGAGAGGTCACAGGCTGACTGACCTGGCTACTCACTTCTTGCTGCACTGACG 1689
QY 41 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
DB 1690 CACTGTGACTTTTATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGT 1749
QY 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTyrPalatnr 80
DB 1750 TACAAGCTATCAAGTCTCAAGTCTCAAGTCTCAAGTCTCAAGTCTCAAGTCTCAAG 1809
QY 81 IleAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
DB 1810 ATTTGGATTGACAGGCTCTCTCTGACTATACCTCTCCAGAAATTTCTTTTGA 1869
QY 101 SerThrThrGlyPheThrLeuTyrGlyMeLeuTyrLysProHisAspLeuGlnProGly 120
DB 1870 AGTACTACTGATTTATACATTTATAGGAGTGTCTCAACACCTCATGATCTACAGCTGGA 1929
QY 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
DB 1930 AAGAATATCTCACTGCTGCTCATATATGGTGTCTCAAGTGTGATGATATAT 1989
QY 134 ----- 134
DB 1990 CGGTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCTCTCTAGCTTATGTGTT 2049
QY 134 ----- 134
DB 2050 GTAAGTATGACACAGGGATCTGTCAACGAGGCTTAAATTTGAAGCGCCTTTAA 2109
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
DB 2110 TATAAAATGGTCAATAATGAAATGACGATCAGGTGGAAGGACTCCAAATATCTAGCTTCT 2169
QY 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyr 171
DB 2170 CGATATGATTTATTCATCTTAGATCTGTGGGCTATCCAGGCTGTCTTATGAGGATTC 2229
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 2230 CTCTCCCTGATGCGATTAATGACAGAGTCAATATCTTCAAGGTTGCTATGCTGGGCTC 2289
QY 132 ProValIleLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
DB 2290 CCAGTCACTCTGTGATCTTCTATGATGATACAGATACAGGAACTTATATGGGTCACTCCT 2349
QY 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
DB 2350 GACCGAATGAAACAGGCTATTAATTAGATCTGTGGCCATGCAAGCAAGAAAGTTCCCTC 2409
QY 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB 2410 TCTGAACCAATCGTTTACTGCTTACATGATGTTCTCTGATGAGAAATTCATTTTGA 2469
QY 253 HistThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
DB 2470 CATACCAAGATATATCTGAGTCTTTTATGACAGGCTGGAAGAAACCAATATGATTTACAGATC 2529
QY 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGlyLysLeuHis 291
```

```
DB 2530 TATCTCAGAGACACACACGCTAAAGATTCTCTGATCTCGAACAATTATGACTGCAT 2589
QY 292 LeuLeuHisTyrLeuGlnGluLysAsnLeuGlySerArgIleAlaAlaLeuLysValIle 310
DB 2590 CTTTGCACTACTCTTCAAGAAACCTTGATCAGCTATTTGCTCTTAAAGTGAATA 2646

RESULT 2
US-10-170-789-39
; Sequence 39, Application US/10170789
; Publication No. US20030180930A1
GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Oladot, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Welch, Nadine
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
FILE REFERENCE: 10448-191001
CURRENT APPLICATION NUMBER: US/10/170,789
CURRENT FILING DATE: 2002-06-13
PRIORITY APPLICATION NUMBER: US 09/797,039
PRIORITY FILING DATE: 2001-02-28
PRIORITY APPLICATION NUMBER: PCT/US01/06525
PRIORITY FILING DATE: 2001-02-28
PRIORITY APPLICATION NUMBER: US 60/186,061
PRIORITY FILING DATE: 2000-02-29
PRIORITY APPLICATION NUMBER: US 09/882,166
PRIORITY FILING DATE: 2001-06-15
PRIORITY APPLICATION NUMBER: PCT/US01/19269
PRIORITY FILING DATE: 2001-06-15
PRIORITY APPLICATION NUMBER: US 60/212,078
PRIORITY FILING DATE: 2000-06-15
PRIORITY APPLICATION NUMBER: US 09/934,406
PRIORITY FILING DATE: 2001-08-21
PRIORITY APPLICATION NUMBER: PCT/US01/26052
PRIORITY FILING DATE: 2001-08-21
PRIORITY APPLICATION NUMBER: US 60/226,740
PRIORITY FILING DATE: 2000-08-21
PRIORITY APPLICATION NUMBER: US 09/861,801
PRIORITY FILING DATE: 2001-05-21
PRIORITY APPLICATION NUMBER: PCT/US01/16549
PRIORITY FILING DATE: 2001-05-21
PRIORITY APPLICATION NUMBER: US 60/205,508
PRIORITY FILING DATE: 2000-05-19
PRIORITY APPLICATION NUMBER: US 09/801,267
PRIORITY FILING DATE: 2001-03-06
PRIORITY APPLICATION NUMBER: PCT/US01/07138
PRIORITY FILING DATE: 2001-03-05
PRIORITY APPLICATION NUMBER: US 60/187,454
PRIORITY FILING DATE: 2000-03-07
PRIORITY APPLICATION NUMBER: US 09/829,671
PRIORITY FILING DATE: 2001-04-10
PRIORITY APPLICATION NUMBER: PCT/US01/40483
PRIORITY FILING DATE: 2001-04-11
PRIORITY APPLICATION NUMBER: US 60/197,508
PRIORITY FILING DATE: 2000-04-18
PRIORITY APPLICATION NUMBER: US 09/961,721
PRIORITY FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: PCT/US01/29904
PRIORITY FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: US 60/235,023
PRIORITY FILING DATE: 2000-09-25
PRIORITY APPLICATION NUMBER: US 10/045,367
PRIORITY FILING DATE: 2001-11-07
PRIORITY APPLICATION NUMBER: US 60/246,561
PRIORITY FILING DATE: 2000-11-07
PRIORITY APPLICATION NUMBER: US 09/801,275
PRIORITY FILING DATE: 2001-03-06
PRIORITY APPLICATION NUMBER: PCT/US01/07074
PRIORITY FILING DATE: 2001-03-05
PRIORITY APPLICATION NUMBER: US 60/187,420
```

PRIOR FILING DATE: 2000-03-07
NUMBER OF SEQ ID NOS: 63
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 2649
TYPE: DNA
ORGANISM: Homo sapiens
US-10-170-789-39

Alignment Scores:

Pred. No.:	1,47e-209	Length:	2649
Score:	1645.50	Matches:	310
Percent Similarity:	86.35%	Conservative:	0
Best Local Similarity:	86.35%	Mismatches:	0
Query Match:	97.95%	Indels:	49
DB:	12	Gaps:	1

US-10-070-464-3 (1-310) x US-10-170-789-39 (1-2649)

QY 1 PheGluGlyThrIysAspSerProLeuGluHisIleuTyValIleSerTyValAsn 20
DB 1570 TTGGAAGGACCAAGAGCTCCCTTTAGAGCATCAGCTGACGTAAGTCACTTAAT 1629
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln 40
DB 1630 CCGAGAGAGCTGACCAAGCTGAGTACCGGTGCTACTCAGATTCCTGCTGATCAGT 1689
QY 41 HisCysAspPhePheIleSerIysTyrSerAsnGlnIysAsnProHisCysValSerLeu 60
DB 1690 CACTGACATCTTTTAAGTAAGTATGTAACAGAAAGATCCACACTGTGTCCCTT 1749
QY 61 TyrLeuLeuSerSerProGluAspAspProThrCysIleThrIysGluPheTPAlaThr 80
DB 1750 TACAGGTATCAAGTCTGTAAGATGACCAACTTGCAAAACAAAGATTTTGGCCACC 1809
QY 81 IleLeuAspSerIleGlyProLeuProAspTyrThrProGluIlePheSerPheGln 100
DB 1810 ATTTTGATTCACAGAGCTCTCTCTCTGACTATACCTCCAGAAATTTCTCTTTGAA 1869
QY 101 SerThrThyGlyPheThrIleuTyrgIyMeIleuTyrgIyPheProHisAspLeuGlnProGly 120
DB 1870 AGTACTGATCGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
QY 121 LysIleTyrgIyProThrValIleuPheIleTyrgIyGlyProGln----- 134
DB 1930 AAGAAATATCTCTAGTGTGCTGATATATGATGATGATGATGATGATGATGATGATGAT 1989
QY 134 ----- 134
DB 1990 CGGTTAAAGAGTCAAGTATTCGCTGATATACCTGACTCTCTAGGTATGTGTT 2049
QY 134 ----- 134
DB 2050 GTAGTATGACAAACAGGAGTCTGTACCGAGGCTTAATTTGAAGCGGCTTTAA 2109
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrgIyLeuAlaSer 151
DB 2110 TATAAATATGCTCAATATGAAATTTGACATCAGGTGAAAGATCTCAATATCTAGCTTCT 2169
QY 152 ArgTyrgAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrgIyGlyTyr 171
DB 2170 CGATATGATTTCACTTACCTAGATCGGTGCGATCCACCGGTGCTGCTTATGAGAGTAC 2229
QY 172 LeuSerLeuMerIleLeuMetGlnIysSerSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 2230 CTCTCCCTGATGCAATTAATGACAGAGTCCAGATATCTTCAGGGGTGCTATGCTGGGGCC 2289
QY 192 ProValThrLeuTrpIlePheTyrgAspThrGlyTyrThrGluArgTyrgIyMetGlyHisPro 211
DB 2290 CCAAGTCACTCTGTGATCTCTCTATGATACAGGATACCGGAAGTATATATGATGATGAT 2349
QY 212 AspGlnAspGluGlnIleTyrgIyLeuGlySerValAlaMetGlnAlaGlyIysPhePro 231

RESULT 3

US-09-976-674-2
Sequence 2, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinaanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 2671
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-2

Alignment Scores:

Pred. No.:	1,49e-209	Length:	2671
Score:	1645.50	Matches:	310
Percent Similarity:	86.35%	Conservative:	0
Best Local Similarity:	86.35%	Mismatches:	0
Query Match:	97.95%	Indels:	49
DB:	9	Gaps:	1

US-10-070-464-3 (1-310) x US-09-976-674-2 (1-2671)

QY 1 PheGluGlyThrIysAspSerProLeuGluHisIleuTyValIleSerTyValAsn 20
DB 1577 TTGGAAGGACCAAGAGCTCCCTTTAGAGCATCAGCTGACGTAAGTCACTTAAT 1636
QY 21 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlyIleSerGln 40
DB 1637 CCGAGAGAGTGAAGAGCTGAGTACCGGTGCTACTCAGATTCCTGCTGATCAGT 1696
QY 41 HisCysAspPhePheIleSerIysTyrSerAsnGlnIysAsnProHisCysValSerLeu 60
DB 1697 CACTGACATCTCTTTTAAGTAAGTATGTAACCAAGAAATCCACACTGTGTCTCTT 1756
QY 61 TyrLeuLeuSerSerProGluAspAspProThrCysIleThrIysGluPheTPAlaThr 80
DB 1757 TACAACTATCAAGTCTTAAGATGACCAACTTGCAAAACAAAGAAATTTTGGGCCACC 1816
QY 81 IleLeuAspSerIleGlyProLeuProAspTyrThrProGluIlePheSerPheGln 100
DB 1817 ATTTTGATTCACAGAGTCTCTCTCTGATATATGATGATGATGATGATGATGATGAT 1876
QY 101 SerThrThyGlyPheThrIleuTyrgIyMeIleuTyrgIyPheProHisAspLeuGlnProGly 120

Db 1877 AGTACTACTGAGTTTACATTTGATGGATGCTCTACAGCCTCATGATCTACAGCCTGGA 1936
Qy 121 LysLysLysProThrValLeuPheLeuGlyGlyProGln----- 134
Db 1937 AAGAAATATCTTACTGCTGCTTATATATGCTGCTTCTACAGGTCAGTTGGTAATAT 1996
Qy 134 ----- 134
Db 1997 CGATTTAAGAGTCAGATATTCCTGCTGAATACCTTACCTCTAGGTTATGTTT 2056
Qy 134 ----- 134
Db 2057 GTAGTATAGACAAACAGGAGTCTCTACCCAGGAGCTTAATTTGAAGCCCTTTAA 2116
Qy 135 -----GlyGlnLeuGluLeuAspGlnValGluGlyLeuGlnValLeuAsp 151
Db 2117 TATATAATGGGTCAATAGAAATGAGATCAGTGGAGAAAGATCCCAATCTTACGTTCT 2176
Qy 152 ArgTyrAspPheLeuAspLeuAspArgValGlyLeuGlyTyrSerTyrGlyGlyTyr 171
Db 2177 CGATATGATTTTCAATGATCTAGATCGTGGGCAATCCAGCGCTGGTCTTATGGAGATAC 2236
Qy 172 LeuSerLeuMetLeuLeuMetGlnArgSerAspLeuPheArgValAlaLeuAlaGlyVal 191
Db 2237 CTCTCCCTGATGGCAATTAATGACAGAGTCAGATATCTTCAAGGTTGCTTATGCTGGAGCC 2296
Qy 192 ProValThrLeuTyrPheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
Db 2297 CCAATGATCTGTCGATCTTCTATGATACAGGATACAGGAAAGCTTATGGGTCACCT 2356
Qy 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
Db 2357 GACAGAGATGAACAGGAGCTATTAATGATGATCTGTGGCAATGCAAGAGAAAGTTCCCC 2416
Qy 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
Db 2417 TCTGAACCAATGCTTACTGCTCTTACATGCTTCTTCTGATGAGAAATGCTCATTTTGA 2476
Qy 252 HisThrSerLeuLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnLe 271
Db 2477 CATACAGATATATTAATGATGATTTTATGATGAGGCTGGAAGCCATATGATTAAGATC 2536
Qy 272 TyrProGlnGluArgHisSerLeuArgValProGluSerGlyGluHisTyrGluLeuHis 291
Db 2537 TATCCCTCAGAGACACACATTAAGATTCCTGAATCGGAGAACATTTATGAACGTCAT 2596
Qy 292 LeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgLeuAlaLeuLeuValLeu 310
Db 2597 CTTTGCACTACCTTCAAGAAACCTTGGATCAGTATTTGCTGCTTAAGAGTATA 2653

RESULT 4
US-10-170-789-37
; Sequence 37, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Weich, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
; TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-191001
; CURRENT APPLICATION NUMBER: US/10/170,789
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797,039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882,166

; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,078
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/934,406
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26052
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (229)...(2874)
US-10-170-789-37
Alignment Scores:
Pred. No.: 1,94e-209 Length: 3143
Score: 1645.50 Matches: 310
Percent Similarity: 86.35% Conservative: 0
Best Local Similarity: 86.35% Mismatches: 0
Query Match: 97.95% Indels: 49
DB: 12 Gaps: 1
US-10-070-464-3 (1-310) x US-10-170-789-37 (1-3143)
Qy 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyrValValSerTyrValAsn 20
Db 1798 TTTGAAGGACCCAAAGATCTCCCTTTAGAGCATCAGTCTAGTACGTTACGTAAT 1857
Qy 21 ProGluValValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlaSerGln 40
Db 1858 CCTGAGAGAGTGACAAAGGCTGAGTACCGGTCTACTCATCATTTGCTGACATCACTGAC 1917

Qy 41 HisCysAspPhePhe11eSerLyTySerAengInLyAsnProHiCysValSerLeu 60
Db 1918 CACTGTGACTTCTTATAGTAGTAGTAACGAAAGAAATCCACACTGTGTCCCTT 1977
Qy 61 TyTyLySerSerSerProGluAspAspProThrCysLySerThyLeGluPheTpa1aThr 80
Db 1978 TACAAGGCTATCAAGCTCGAAGATGACCACTTCGAAACAAAGAAATTTGGGCCACC 2037
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyThrThrProProGlu11ePheSerPheGlu 100
Db 2038 ATTGTGGATTCAGCAGGCTCTCTCTCTAGCTACTCTCCAGAAATTTCTCTTTTGA 2097
Qy 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyTySerProHiAspLeuGlnProGly 120
Db 2098 AGTACTACTGGATTACATTGATGATGCTCTCAAGCCCTCATGATCTACAGCCTGGA 2157
Qy 121 LysLySerTyProThrValLeuPhe11eTyGlyGlyProGln----- 134
Db 2158 AAGAAATATCTCTACTGTCTGTCTATATATGATGCTCTCAGGTGCACTGGTGAATAT 2217
Qy 134 ----- 134
Db 2218 CGGTTTAAAGATCAAGTATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGTT 2277
Qy 134 ----- 134
Db 2278 GTAGTAGTAGACAACAGGGGATCTCTGACCGAGGGCTTAATTGAAGGGCCTTAA 2337
Qy 135 -----GlyGln11eGlu11eAspAspGlnValGluGlyLeuGlnTyTyLeuAlaSer 151
Db 1338 TATATAATGGGTCAAAATGGAATGACGATCAGGTGGAAGGATCTCAATATCTAGCTCT 2397
Qy 152 ArgTyAspPhe11eAspLeuAspArgValGly11eHISGlyTyTySerTyGlyGlyTy 171
Db 2398 CGATATGATTTTCTTACTTATAGATGCTGTGGCATCCAGCGGTGCTCTATGAGGAATAC 2457
Qy 172 LeuSerLeuMetAlaLeuMetGlnAspSerAsp11ePheArgValAla11eAlaGlyAla 191
Db 2458 CTCCTCCGATGGCATTAATGATGACAGAGGTCAATATCTTCAAGGTTGCTATGGGGCC 2517
Qy 192 ProValThrLeuThrP11ePheTyAspThrGlyTyThrGluAspGlyMetGlyHisPro 211
Db 2518 CCACTACTCTGTGTGATCTTCTATGATGATCAGGATACAGGAACGTTATATGGGTACCCCT 2577
Qy 212 AspGlnAsnGlnGlnGlyTyTyLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
Db 2578 GACCAAGATGAAACAGGGCTATTACTTAGGATCTGTGGCATGCAAGCAAGAAAGTTCCCC 2637
Qy 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
Db 2638 TCTGAACCAATTCGTTTACTGCTCTTACATGCTTCCGATGAGAAATGCTCATTTTGA 2697
Qy 252 HisThrSer11eLeuLeuSerPheLeuVal11eAlaGlyLysProTyAspLeuGln11e 271
Db 2698 CATACCAATATATATGATGATTTTACTGAGGGCTGGAAGCCATATGATTTACAGATC 2757
Qy 272 TyTyProGlnGluAspHisSer11eArgValProGluSerGlyGlnHisTyTyGluLeuHis 291
Db 2758 TATCTCAAGGAGAACACAGCATTAAGATTCCTTAATCTGGGAGCAATTTAAGAACTGAT 2817
Qy 292 LeuLeuHisTyTyLeuGlnGluAsnLeuGlySerArg11eAlaAlaLeuVal11e 310
Db 2818 CTTTTCACACTTACTTCAAGAAACCTTGATCAGTATGCTGCTTAAAGTATATA 2874

APPLICANT: Junten, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPIP
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976,674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 4676
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-20
Alignment Scores:
Pred. No.: 5,05e-209 Length: 4676
Score: 1644.50 Matches: 308
Percent Similarity: 99.35% Conservative: 0
Best Local Similarity: 99.35% Mismatches: 1
Query Match: 97.89% Indels: 2
DB: 10 Gaps: 1
US-10-070-464-3 (1-310) x US-09-976-674-20 (1-4676)
Qy 1 PheGluGlyThrThyAspSerProLeuGlnHisHisLeuTyValValSerTyValAsn 20
Db 1783 TTTGAAGCACCAAGACTCCCTTTAGAGCATACCTGTGAGTAGTACGTAAT 1842
Qy 21 ProGlyGluVal11eThrArgLeuThrAspArgGlyTyTySerHisSerCysLe11eSerGln 40
Db 1843 CCGGAGAGGAGACAGGCTGACGACCGGTGCTACTACATTTCTGTGATCGTCAAG 1902
Qy 41 HisCysAspPhePhe11eSerLyTySerAengInLyAsnProHiCysValSerLeu 60
Db 1903 CACTGTGACTTCTTATAGTATAGTAAAGTAAAGAAATCCACACTGTGTCTCCCTT 1962
Qy 61 TyTyLySerSerSerProGluAspAspProThrCysLySerThyLeGluPheTpa1aThr 80
Db 1963 TACAAGGCTATCAAGCTCTGAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACC 2022
Qy 81 IleLeuAspSerAlaGlyProLeuProAspTyThrThrProProGlu11ePheSerPheGlu 100
Db 2023 ATTTGGATTCAGCAGGCTCTCTCTGACATATCTCTCAGAAATTTCTCTTTGAA 2082
Qy 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyTySerProHiAspLeuGlnProGly 120
Db 2083 AGTACTACTGGATTACATTGTATGGGATGCTTACAAAGCCTCATGATCTACAGCCTGA 2142
Qy 121 LysLySerTyProThrValLeuPhe11eTyGlyGlyProGlnGlyGln11eGlu11eAsp 140
Db 2143 AAGAAATATCTCTACTGTGCTGTCTATATGTGTGTGCTG---GGTCAAAATGAAATTAAC 2198
Qy 141 AspGlnValGluGlyLeuGlnTyTyLeuAlaSerArgTyAspPhe11eAspLeuAspArg 160
Db 2199 GATAGGTGAGAAAGACTTCAATATCTAGCTTCTGATATGATTTTCATTGACCTTATGAT 2258
Qy 161 ValGly11eHisGlyTyTySerTyGlyGlyTyTyLeuSerLeuMetAlaLeuMetGlnArg 180
Db 2259 GTGGGATCCACAGGCTGTCTTATGAGATACCTCTCCGATGGCATTAATGACAGAGG 2318
Qy 181 SerAsp11ePheAspValAla11eAlaGlyAlaProValThrLeuThrP11ePheTyAsp 200
Db 2319 TCGATATCTTCAGGGTGTGCTATGTGGGGCCCACTGATCTGTGATCTTCTATGAT 2378
Qy 201 ThrGlyTyThrGluArgTyMetGlyHisProAspGlnAsnGlnGlnGlyTyTyLeu 220
Db 2379 ACAAGATACACGGAAGCTTATATGGGTCACTTGAACAGAAATGAACAGGCTATTACTTA 2438
Qy 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240
Db 2439 GATCTGTGTCATGCAAGCAAGAAAGTTCCCTTGAAACCAATCGTTACTGCTCTTA 2498

RESULT 5
US-09-976-674-20
Sequence 20, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre

QY 241 HisGlyPheLeuAspGluAsnValHisPheAlaHisSerSerLeuLeuSerPheLeu 260
DB 2499 CATGTTTCCTCGATGAGAAATGTCATTTTGACATACAGATATTAAGTACTGAGTTTTTA 2558
QY 261 ValAlaGlaGlyLysProThrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280
DB 2559 GTGAGGGCTGGAGAAAGCCATATGATTTTACAGATCTATCTCAGAGAAAGACACAGCATTAAGA 2618
QY 281 ValProGluSerGlyGluHisTyrGluLeuHisIleTyrLeuGlnGluAsnLeu 300
DB 2619 GTTCTGTAATCGGAGAAACATTAATGAACTGCATCTTTTGCACTACTTCAAGAAAACCTT 2678
QY 301 GlySerArgIleAlaAlaLeuLysValIle 310
DB 2679 GGATCAGCATTTGCTGCTCTTAAGAGTATA 2708
RESULT 6
US-09-976-674-14
; Sequence 14, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 4309
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-14
Alignment Scores:
Pred. No.: 1,566-206 Length: 4309
Score: 1625.50 Matches: 310
Percent Similarity: 85.87% Conservative: 0
Best Local Similarity: 85.87% Mismatches: 0
Query Match: 96.76% Indels: 51
Gaps: 1
US-10-070-464-3 (1-310) x US-09-976-674-14 (1-4309)
QY 1 PheGluGlyThrLysAspSerProLeuGluHisIleLeuTyrValValSerTyrValAsn 20
DB 1263 TTTGAAGGCACCAAGACTCCCTTTAGAGCATCACTTACGTAGTACGTTAAAT 1322
QY 21 ProGluGluValThrArgLeuThrAspArgLysTyrSerHisSerCysValSerGln 40
DB 1323 CTTGGAGAGTACCAAGGCTGAGCTGACCTGGCTACTCATCTTCTGCAATCGATCAG 1382
QY 41 HisCysAspPhePheIleSerLysTyrSerArgGlnLysAsnProHisCysValSerLeu 60
DB 1383 CACTGTGACTCTTTATATAGTAAGTATAGTAACCAAGAAATCCACAGCTGTGTCCTT 1442
QY 61 TyrIleValSerSerProGluLysAspProThrCysIleThrLysGluPheTrpAlaThr 80
DB 1443 TACAAGCTATCAAGCTCGAAGATGACCCAACTTGCAAAAAGAAATTTGGGCGCAC 1502
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluGlnIlePheSerPheGlu 100
DB 1503 ATTTGGATTCACGAGTCTCTCTCTGACTATACCTCCCAAGAAATTTCTTTTGA 1562
QY 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
DB 1563 AGTACTACTGATTAATTAATTTATGAGATGCTTACAAAGCCTCATGATCTACAGCCTGGA 1622

QY 121 LysIleTyrProThrValLeuPheIleTyrGlyGly--ProGln----- 134
DB 1623 AAGAAATATCTACTGTCTGTCTATATATGTTGGTCTCTCAGGTGAGTTGTGAATA 1682
QY 134 ----- 134
DB 1683 ATCGGTTAAAGAGTCAAGTATTTCCGCTTGAAATACCTTAGCTCTTAGTTATGTGG 1742
QY 134 ----- 134
DB 1743 TTGTAGTATAGACAACAGGGATCTGTCAACAGGCTTAAATTTGAAGCGCTTTA 1802
QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 151
DB 1803 AATATTAATAGGTCAATTAAGAAATTTGACGATCAGGTGGAAGAACTCCAAATATCTAGCTT 1862
QY 151 GATAGTAAAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyrGlyT 171
DB 1863 CTGATATGATTTTCACTTGAATCGTGTGGGATCCAGGCTGAGCTGTCCTATGAGAGAT 1922
QY 171 YrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 191
DB 1923 AACTTCCCTGATGSCATTAATGACAGCTCAGATATCTTCAGGCTTCTATGCTGGGG 1982
QY 191 IAProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisP 211
DB 1983 CCCAGTCACTCTGGATCTTCTATGATGATACAGGATACAGGAAGTTATATGGGCTCAC 2042
QY 211 RoAspGluAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPheP 231
DB 2043 CTGACCAAGATGAACAGGCTTATTACTTAAGATCTGTGCCATGACGACAGAAAGTTCC 2102
QY 231 RoSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251
DB 2103 CCTCGAACCAATGCTTACTGCTCTTACATGCTTCTCGATAGAAATGCCATTTTG 2162
QY 251 IAspThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnI 271
DB 2163 CACATACCAATATATTAATGAGTTTATGAGAGGCTGGAAGCCATATGATTTTACAGA 2222
QY 271 IeTyrProGluGluArgHisSerIleArgValProGluSerGlyGluHisTyrGlyLeuH 291
DB 2223 TCTATCTCAAGAGACACAGCATTAAGTCTGAAATCGGAGAACATTAATGAACTGC 2282
QY 291 IeLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLysValIle 310
DB 2283 ATCTTTGCATCACTCTTCAAGAAAACCTTGATCACTGATTGCTGCTCAAAAAGTGATA 2341
RESULT 7
US-09-976-674-12
; Sequence 12, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-12
Alignment Scores:
Pred. No.: 1,886-206 Length: 4829

Score: 1625.50 Matches: 310
 Percent Similarity: 85.87% Conservative: 0
 Best Local Similarity: 85.87% Mismatches: 0
 Query Match: 96.76% Indels: 51
 DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x US-09-976-674-12 (1-4829)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyValIleSerTyValAsn 20
 DB 1783 TTGGAAGGACCAAGACTCCCTTTAGACATCACTGACGTACGTACGTAAAT 1842
 QY 21 ProGlyValValThrArgLeuThrAspArgIleTySerHisSerCysCysIleSerGln 40
 DB 1843 CCGAGAGAGGTGACAAAGGCTGACGACCGGTGCTACTACATCTTCTGCTGACAGCAG 1902
 QY 41 HisCysAspPhePheIleSerIleTySerSerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1903 CACTGTCACCTCTTTATTAAGTATGATTAACCAAGAAATCCACACTGCTGCTCTT 1962
 QY 61 TyLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80
 DB 1963 TACAGCTATCAAGTCTGTAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACC 2022
 QY 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGlu 100
 DB 2023 ATTTGGATTACAGAGTCTCTCTTCCGACTATACCTCCCAAGAAATTTCTCTTTGAA 2082
 QY 101 SerThrThrGlyPheThrLeuTyArgLysLeuTyLysProHisAspLeuGlnProGly 120
 DB 2083 AGTACTACTGATTTATCATTTATGAGGATGCTCTACAAAGCTCATGATCTACAGCCCTGA 2142
 QY 121 LysLysTyProThrValLeuPheIleTyGlyGly--ProGln----- 134
 DB 2143 AAGAAATATCTACTGCTGCTGCTCATATATGATGAGTGCTCTCTCAGGTGCACTTGAGATA 2202
 QY 134 ----- 134
 DB 2203 ATCGGTTTAAAGAGTCAATATTCCGCTTGAATACCTAGCCTCTCATGATTATGTGG 2262
 QY 134 ----- 134
 DB 2263 TTGTAGTATGATGACAAACAGGGAGTCTCTGTCACCGAGGCTTAAATTGAGGCCCTTTA 2322
 QY 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAsn 151
 DB 2323 AATATATAATGGGTCAATATGAATATGACATGAGGTGAGAAAGACTCCCAATATCTAGCTT 2382
 QY 151 exArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyGlyTyr 171
 DB 2383 CTGCAATATGATTTCACTTGAATGATGCTGGGCAATCCAGGCTGCTCTATGAGAGAT 2442
 QY 171 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 191
 DB 2443 ACCTCTCCCTCGATGAGCATTTATGACAGAGTCAATATCTTCAAGGTTGCTATGCTGGGG 2502
 QY 191 LeuProValThrLeuTyrPheTyAspThrGlyTyrThrGlnLysArgTyMetGlnHisP 211
 DB 2503 CCCCAAGTCACTGCTGATTTCTATGATACAGAGATACAGAGAACTTATATGAGGTCAAC 2562
 QY 211 roAspGlnAsnGlnGlnGlyTyrTyLeuGlySerValAlaMetGlnAlaGluLysPheP 231
 DB 2563 CTGACACAGATGAAAGGGCTATTTACTTAGATCTGTGGCCATGCAAGAGAAAGTTC 2622
 QY 231 roSerGluProAsnAlaGluLeuLeuHisGlyPheLeuAspGluAsnValHisPheA 251
 DB 2623 CCTCTAAACCAATTCGTTTACTCTCTTACATGATTTCCGATGAGTGAAGATGCTCATTTTG 2682
 QY 251 LHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnI 271
 DB 2683 CACATACCAATATTAATGAGTTTATGATGAGGGCTGAGAAAGCCATATGATTTACAGA 2742
 QY 271 LeTyProGlnGlnArgHisSerIleArgValProGluSerGlyGluHisTyGlyLeuH 291

DB 2743 TCTATCTCGAGAGACACAGCATTAAGGTTCTGTAATCGGAGAACTTATGAATGC 2802
 QY 291 ILeuLeuHisTyTyLeuGlnGluAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310
 DB 2803 ATCTTTGCACTACTTCAAGAAACCTTGAGTACGATATGCTGCTCTAAAGTGATA 2861

RESULT 8

US-09-976-674-22
 ; Sequence 22, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akimanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Untien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DBPIV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; PRIOR FILING DATE: 2001-10-12
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 22
 ; LENGTH: 4685
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-976-674-22

Alignment Scores:

Pred. No.: 1,176-170 Length: 4685
 Score: 1358.50 Matches: 271
 Percent Similarity: 88.61% Conservative: 9
 Best Local Similarity: 85.76% Mismatches: 27
 Query Match: 80.86% Indels: 9
 DB: 10 Gaps: 5

US-10-070-464-3 (1-310) x US-09-976-674-22 (1-4685)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLeuTyValIleSerTyValAsn 20
 DB 1783 TTGGAAGGACCAAGACTCCCTTTAGACATCACTGACGTACGTACGTAAAT 1842
 QY 21 ProGlyValValThrArgLeuThrAspArgIleTySerHisSerCysCysIleSerGln 40
 DB 1843 CCGAGAGAGGTGACAAAGGCTGACGACCGGTGCTACTACATCTTCTGCTGACAGCAG 1902
 QY 41 HisCysAspPhePheIleSerIleTySerSerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 1903 CACTGTCACCTCTTTATTAAGTATGATTAACCAAGAAATCCACACTGCTGCTCTT 1962
 QY 61 TyLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 80
 DB 1963 TACAAAGTATCAAGTCTGTAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACC 2022
 QY 81 IleLeuAsp----SerAlaGlyProLeuProAspTyThr-----ProProGluIleP 97
 DB 2023 ATTTGGATTACAGTCTCTCAAGTGCATGCTGATGATGATGATGATTAAGAGTCAAGTATT 2082
 QY 97 eSerPheGluSerThrThrGlyPheThrLeuTyArgLysLeuTyLysProHisAsp-- 116
 DB 2083 TCGGCTTGAATACCTCAAGCTCTCTATGATGATGATGATGATGATGATGATGATGAT 2142
 QY 117 -LeuGlnPro-GlyLysTyTyProThrValLeuPheIleTyGlyGlyProGlnGlyG 136
 DB 2143 CCGTACACGAGAGGCTTAAATTGAAAGGCC---TTTAAATAT-----AAAATGGGT 2193
 QY 136 ImlleGluIleAspAspGlnValGluGlyLeuGlnTyLeuAlaSerArgTyAspPheI 156
 DB 2194 AATATGAATTTACAGATGAGGTGAGAAAGACTCCCAATATCTTCTCATATGATTTCA 2253
 QY 156 LeAspLeuAspArgValGlyIleHisGlyTyrPheSerTyGlyGlyTyLeuSerLeuMetA 176

```

Db      2254 TTGACTTAGATCGTGTGGGCATCCACGGCTGTCCTATGAGAGATACCTCCCTGATGG 2313
Qy      176  |aleuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuT 196
Db      2314 CATTAAAGCAGAGTCGATATCTTCAGGGTGTGTATTTGGGGCCCGCATGCTCTGT 2373
Qy      196  |rIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluG 216
Db      2374 GGATCTTCTATGATACGAGTATACCGGACGTTATATAGGCTACCCGACAGATGTAAC 2433
Qy      216  |InGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsn 226
Db      2434 AGGCTATTAAGTATGATCTGTGGCCATGCACAGCAAAAGTTCCCTCTGAACCAATC 2493
Qy      236  |rGluLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIle 256
Db      2494 GTTAACTGCTTATCATGTGTTCTCGATGAGAAATGTCATTTGGACATACCAATATAT 2553
Qy      256  |euleuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGlu 276
Db      2554 TACTGACTTTTATAGTATAGGGCTGGAAAGCCATATGATTTACAGATCTATCTCAGAGA 2613
Qy      276  |rGlySerIleArgValProGluSerGlyGluHisTyrGlyLeuHisLeuLeuHisTyrL 296
Db      2614 GACACAGCATTAAGAGTTCCTGAATCGGAGAAACATATGAACTGCATCTTTGACATACC 2673
Qy      296  |euGlnGluAsnLeuGlySerArgIleAlaIleAlaLeuLysValIle 310
Db      2674 TTCAAGAAACCTTGATCAAGTATCTGCTCTTAAAGTGATA 2717

RESULT 9
US-09-976-674-8 ; Sequence 8, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4523
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-8

Alignment Scores:
Pred. No.:      8.23e-170      Length:      4523
Score:          1352.00      Matches:      257
Percent Similarity: 82.90%      Conservative: 0
Best Local Similarity: 80.48%      Mismatches: 1
Query Match:      80.48%      Indels:      53
DB:              10          Gaps:          1

US-10-070-464-3 (1-310) x US-09-976-674-8 (1-4523)
Qy      1  PheGluGlyThrLysAspSerProLeuGluHisIleSerTyrValIleSerTyrValAsn 20
Db      1783 TTGGAAGGACCAAGACTCCCTTTAGAGCATACCTGTACGTAAGTACAGTAACTAAAT 1842
Qy      21  ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerGlyValIleSerGln 40
Db      1843 CCTGAGAGGCTGACAAAGCTGACGACCGTGGCTACTACATCTTCTGTCATCAGTACAG 1902
Qy      41  HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 60
```

```

Db      1903 CACTGAGACTTCTTTATAGTAAAGTATAGTAACCAAGAAATCACACTGTGTCTCCTT 1962
Qy      64  TyrLysLeuSerSerProGluAspAspProThrCysLysThrTyrGluPheTyrAlaThr 80
Db      1963 TACAACTATCAAGTCTCTGAAGATGACCCAACTTGGAAAAAAGAAATTTTGGCCACC 2022
Qy      81  IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
Db      2023 ATTTTGATTCACAGAGTCTCTCTTCTGACTATACCTCTCCAAATATTTCTCTTTGAA 2082
Qy      101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
Db      2083 AGTACTAGTGAATTTACATGTATGAGGATCTCTACAAAGCTCATGATCTACAGCTGGA 2142
Qy      121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnGlnIleGluIleAsp 140
Db      2143 AAGAAATATCTTACTGTGCTGTTCATATATGTGTGTT----- 2178
Qy      141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160
Db      2178 ----- 2178
Qy      161 ValGlyIleHisGlyTyrPserTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180
Db      2178 ----- 2178
Qy      181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrPheTyrAsp 200
Db      2179 -----CGGTT-GCTATTTCTGGGGCCCGCATCTGTGGATCTTCTATGAT 2225
Qy      201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeu 220
Db      2226 ACAGATATACCGGAAGCTTATATAGGCTACCCGACCAAGATGAACAGGGCTATTACTTA 2285
Qy      221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeu 240
Db      2286 GGATCTGTGGCCATGCAAGCAAGAAAGTTCCCTCTGAACCAATCGTTTACGCTCTTA 2345
Qy      241 HisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 260
Db      2346 CATGGTTCTCTGATGAGAAATGTCATTTGGACATACCAAGTATATTAAGTGAAGTTTATA 2405
Qy      261 ValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArg 280
Db      2406 GTGAGGGCTGGAAGCCATATGATTTTACAGATCTATCTCAGAGACACAGCATAGAGA 2465
Qy      281 ValProGluSerGlyGluHisTyrGlyLeuLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 300
Db      2466 GTTCCTGAATCGGAGAAACATATGAACTGCATCTTTGACATCACTTCAAGAAAACCTT 2525
Qy      301 GlySerArgIleAlaIleAlaLeuLysValIle 310
Db      2526 GGATCAAGTATCTGCTCTTAAAGTGATA 2555

RESULT 10
US-09-976-674-4 ; Sequence 4, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
```

```
/ LENGTH: 2617
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-976-674-4

Alignment Scores:
Pred. No.: 1,59e-135 Length: 2617
Score: 1093.50 Matches: 200
Percent Similarity: 69.34% Conservative: 42
Best Local Similarity: 57.31% Mismatches: 58
Query Match: 65.09% Indels: 49
DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x US-09-976-674-4 (1-2617)

QY 1 PheGluGIYThrLysASPserProLeuGIuHISLeuTYrValISerTYrValaen 20
DB 1553 TTCCAGGGGACCAAGACAGCGCGCTGAGACACACCTTACGTGCTAGTATGAGGGG 1612
QY 21 ProGluGIuValThrArgLeuThrAspArgGlyTYrSerHISerCySeIleserGln 40
DB 1613 GCGGGGAGATGCTACCGCTTCCACACCGCGCTTCCCATAGCTGCTCCATGAGCCAG 1672
QY 41 HISCyASPpPheHISerLysTYrSerAenGlnLYAsnProHISCyValSerLeu 60
DB 1673 AACTTCGACATGCTGCTGACGCCACTACAGCAGCGTGAGCAGCCCGCTGCGTGCACGTC 1732
QY 61 TYrLysLeuSerSerProGluAspAspProThrCYsLYeThrLysGluPheThrAlaThr 80
DB 1733 TACAAAGCTGAGGGGCGCCGACACAGCCCGCTGCACACAGCCCGCTTGGGCTAGC 1792
QY 81 ILeuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePheSerPheGlu 100
DB 1793 ATGATGAGAGCGAGCGAGCGCGCGGATATGTTCTCCAGAGATCTTCATTTCAC 1852
QY 101 SerThrThrglyPheThrLeuTYrGlyMetLeuTYrLYsProHISAspLeuInProGly 120
DB 1853 ACGCGCTCGAGTGTGCGCTTACGCGCATGATCTCAAGCCCGCAGCCTTGACGAGGG 1912
QY 121 LysLYsTYrProThrValLeuPheHISerGlyGlyProGln----- 134
DB 1913 AAGAAGCACCCACCGTCTCTTGTATATGAGGCCCGCCAGGTGACGTGTGATATAC 1972
QY 134 ----- 134
DB 1973 TCCTTCAAGGATCAAGTACTTGCGGCTCAACACACTGCGCTCCCTGGGCTACCGCGTG 2032
QY 134 ----- 134
DB 2033 GTTGTGATTGACGGGAGGGGCTCTGTACAGGAGGCTTCGGTTCAGAGGGGCGCTGA 2092
QY 135 -----GlyGlnIleGlnIleAspAspGlnValGluGlyLeuGlnThyTrieuAlaSer 151
DB 2093 AACCAATGGGCGAGGTGAGATCGAGGACAGGAGGAGGCGCTGAGTTCGTGCGCAG 2152
QY 152 ArgTYrAspPheHISAspLeuAspArgValGlyIleHISGlyTTPSerTYrGlyTYr 171
DB 2153 AAGTATGCTTCATGACCTGAGCGGAGTGCATTCATGCTGATGCTTCACGGGGGCTTC 2212
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 2213 CTCTGCTCATGTGGGCTAATCCACAAAGCCCGAGGTGTTTCAAGTGGCCATCCGGGGTGC 2272
QY 192 ProValIThrLeuThrIlePheTYrAspThrArgLYrThrGluArgTYrMetGlyIlePro 211
DB 2273 CCGGTCACCGTGTGATGCTTACGACACAGGATACATGAGCGCTTACATGAGCGTCCCT 2332
QY 212 AspGlnAenGluGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlyIleAspPro 231
DB 2333 GAGAAACAACAGCAGCAGGCTATGAGCGGTTCCGTGGCCCTGACGTGAGAGAGCTGCC 2392
QY 232 SerGluProAsnArgLeuLeuLeuLeuHISGlyPheLeuAspGluAsnValHISpHeAla 251
```

```
DB 2393 AATGAGCCCAACCGCTTGCTTATCTCCAGGCTTCTGAGGAAAACGTGACATTTTTC 2452
QY 252 HISThrSerIleLeuLeuSerPheLeuValArgAlaGlyLYsProTYrAspLeuGlnIle 271
DB 2453 CACACAACTTCTGCTGCTCCACTGATCCAGACAGGAGAACTTTACAGCTCCAGATC 2512
QY 272 TYrProGluArgHISerIleArgValProGluSerGlyGlyHISerTYrGluLeuHIS 291
DB 2513 TACCCCAACGAGAGACAGTATTTGCTGCGCCCGGAGTCCGGGAGAGACTATGAGTCAG 2572
QY 292 LeuLeuHISerTYrLeuGlnGluAsnLeu 300
DB 2573 TTGCTGACATTTCTACAGGAATATCCTC 2599

RESULT 11
US-09-976-674-28
/ Sequence DB, Application US/09976674
/ Patent No. US20020115843A1
/ GENERAL INFORMATION:
/ APPLICANT: Qi, Steve
/ APPLICANT: Akinsanya, Karen
/ APPLICANT: Riviere, Pierre
/ APPLICANT: Jundten, Jean-Louis
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DDPV
/ FILE REFERENCE: 70669
/ CURRENT APPLICATION NUMBER: US/09/976,674
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,117
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 28
/ LENGTH: 4219
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-976-674-28

Alignment Scores:
Pred. No.: 3,47e-135 Length: 4219
Score: 1093.50 Matches: 200
Percent Similarity: 69.34% Conservative: 42
Best Local Similarity: 57.31% Mismatches: 58
Query Match: 65.09% Indels: 49
DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x US-09-976-674-28 (1-4219)

QY 1 PheGluGIYThrLysASPserProLeuGIuHISLeuTYrValISerTYrValaen 20
DB 1909 TTCCAGGGGACCAAGACAGCGCGCTGAGACACACCTTACGTGCTAGTATGAGGGG 1968
QY 21 ProGluGIuValThrArgLeuThrAspArgGlyTYrSerHISerCySeIleserGln 40
DB 1969 GCGGGGAGATGCTGAGCTTCCACACCGCGCTTCCCATAGCTGCTCCATGAGCCAG 2028
QY 41 HISCyASPpPheHISerLysTYrSerAenGlnLYAsnProHISCyValSerLeu 60
DB 2029 AACTTCGACATGCTGCTGACGCCACTACAGCAGCGTGAGCAGCCCGCTGCGTGCACGTC 2088
QY 61 TYrLysLeuSerSerProGluAspAspProThrCYsLYeThrLysGluPheThrAlaThr 80
DB 2089 TACAAAGCTGAGGGGCGCCGACGACAGCCCGCTGCACAAAGCAGCCCGCTTGGGCTAGC 2148
QY 81 ILeuAspSerAlaGlyProLeuProAspTYrThrProProGluIlePheSerPheGlu 100
DB 2149 ATGATGAGAGGAGCGAGCTGCGCGCGGATATGTTCTCCAGAGATCTTCATTTCAC 2208
QY 101 SerThrThrglyPheThrLeuTYrGlyMetLeuTYrLYsProHISAspLeuInProGly 120
DB 2209 ACGCGCTGAGTGTGCGCTTACGCGCATGATCTCAAGCCCGCAGCCTTGACGAGGG 2268
QY 121 LysLYsTYrProThrValLeuPheHISerGlyGlyProGln----- 134
```


Db 2269 AAGAGACACCCACCGCTCTTGTATATGAGAGCCGCCAGGTGACGTGTGATATAC 2328
 134 ----- 134
 Db 2329 TCCTTCAAGGATCAAGTACTGGGCTCAACACACTGGCCTCCCTGGGCTAACGCCGTG 2388
 134 ----- 134
 Db 2389 GTTGTGATTGACGGCAGGGGCTCTGTACGACGAGGCTTCGGTTGGAAGGGGCCCTGAAA 2448
 135 -----GlyGlnIleGluIleAspArgIleValIleGlyLeuGlnIleAspArg 151
 2449 AACCAATATGGCCAGTGAAGATCGAGACCAAGTGGAGGGCTGTGAGTTCCGTGGCCGAG 2508
 152 ArgTyrAspPheIleAspLeuAspArgValIleGlyIleAspArgTyrGlyIleTyr 171
 2509 AAGTATGGCTTACATGACCTGAGCCAGATGGTCCATCAGCTGAGTGTCTTACGGGGCTTC 2568
 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 2569 CTCTGGCTCATGGGCTTATTCACCAAGCCCAAGGTTTCAGGTGGCCATCCGGGGTCC 2628
 192 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 2629 CCGGTACCGGTGTGATGGCTTACGACAGAGGTATACGAGCGCTACAGACGTTCAC 2688
 212 AspGlnAspGlnIleGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlnIleAspPro 231
 2689 GAGAACACACACGACGAGCTATAGAGGGGTTCCGTGGCCCTGACCTGAGAACCTGCCCT 2748
 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAla 251
 2749 AATGAGCCCAACCGCTTGTCTTATCCCAAGGCTTCTGACGAAAACTGCACCTTTTC 2808
 252 HistTrpSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
 2809 CACACAACTTCTGCTCTCCCACTGATCCAGCGGAGAACTTACCAAGCTCCAGATC 2868
 272 TyrProGlnIleArgHisSerIleArgValProGluSerGlyGlnIleTyrGlnLeuHis 291
 2869 TACCCCAAGAGACAGACAGTATTCGCTCCCGAGTCCGGGAGAGCACTATGAAGTCACG 2928
 292 LeuLeuHisTyrLeuGlnIleAsnLeu 300
 2929 TTGCTGCACTTCTTACAGGAATACCTC 2955
 RESULT 12
 US-09-976-674-24
 ; Sequence 24, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akinsanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; PRIOR FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 24
 ; LENGTH: 4302
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-976-674-24
 Alignment Scores:
 Pred. No.: 3,59e-135 Length: 4302
 Score: 1093.50 Matches: 200
 Percent Similarity: 69.34% Conservative: 42

Best Local Similarity: 57.31% Mismatches: 58
 Query Match: 65.09% Indels: 49
 DB: 10 Gaps: 1
 US-10-070-464-3 (1-310) x US-09-976-674-24 (1-4302)
 1 PheGluGlyThrLysAspSerProLeuGlnIleHisLeuTyrValIleSerTyrValAsn 20
 1909 TTCAGGGGACCAAGACACCGCGCTGAGACCAACCTTACGTGTGACATGAGCGG 1968
 21 ProGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysAlleSerGln 40
 1969 GCCGCGAGATGTGACGCTTCCACACGCGCGCTTCTCCCAATAGCTGTCCATGAGCCG 2028
 41 HisCysAspPhePheIleSerTyrSerAspGlnIleAspProHisCysValSerLeu 60
 2029 AACTTGACATGTGTGTCAGCCATACACAGACGAGAGACCGCCGCTTGGTGCACGTC 2088
 61 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheThrAlaThr 80
 2089 TACAAAGTGAAGCGGCCGACAGACACCCCTGCACAAAGACCCCGCTTGGGCTTAC 2148
 81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
 2149 ATGATGAGGACGACGACGCTGCCCGCGATATATGTTCTCCAGAGATCTTCCATTCCAC 2208
 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 120
 2209 ACGCGCTGAGATGTGCGGCTTACGACATGATCTACAAAGCCCAAGCCCTTGCAGCCAGG 2268
 121 LysLysTyrProThrValLeuPheIleTyrGlyIleProGln----- 134
 2269 AAGAGACACCCACCGCTCTTGTATATGAGAGCCGCCAGGTGACGTGTGATATAC 2328
 134 ----- 134
 2329 TCCTTCAAGGATCAAGTACTGGGCTCAACACACTGGCCTCCCTGGGCTAACGCCGTG 2388
 134 ----- 134
 2389 GTTGTGATTGACGGCAGGGGCTCTGTACGACGAGGCTTCGGTTGGAAGGGGCCCTGAAA 2448
 135 -----GlyGlnIleGluIleAspArgIleValIleGlyLeuGlnIleAspArg 151
 2449 AACCAATATGGCCAGTGAAGATCGAGACCAAGTGGAGGGCTGACGTGTGAGCGGAG 2508
 152 ArgTyrAspPheIleAspLeuAspArgValIleGlyIleAspArgTyrGlyIleTyr 171
 2509 AAGTATGGCTTACATGACCTGAGCCAGATGGTCCATCAGCTGAGTGTCTTACCGGGGCTTC 2568
 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 2569 CTCTGGCTCATGGGCTTATTCACCAAGCCCAAGGTTTCAGGTGGCCATCCGGGGTCC 2628
 192 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 2629 CCGGTACCGGTGTGATGGCTTACGACAGAGGTATACGAGCGCTACAGACGTTCAC 2688
 212 AspGlnAspGlnIleGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlnIleAspPro 231
 2689 GAGAACACACGACGAGCTATAGAGGGGTTCCGTGGCCCTGACCTGAGAACCTGCCCT 2748
 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAla 251
 2749 AATGAGCCCAACCGCTTGTCTTATCCCAAGGCTTCTGACGAAAACTGCACCTTTTC 2808
 252 HistTrpSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIle 271
 2809 CACACAACTTCTGCTCTCCCACTGATCCGAGCGGAGAACTTACCAAGCTCCAGATC 2868
 272 TyrProGlnIleArgHisSerIleArgValProGluSerGlyGlnIleTyrGlnLeuHis 291
 2869 TACCCCAAGAGACAGACAGTATTCGCTCCCGAGTCCGGGAGAGCACTATGAAGTCACG 2928


```

QY      22  glygluvalthrargleuthrhaparglytyrserhissercysalserglnhis 41
      1933  GGGGAGATGCTAGCGCTCACACGCGGCTCTCCATAGCTGCTCCATGAGCCAGAAC 1992
QY      42  CysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisCysValSerLeuTYr 61
      1993  TTCGACATGTCCTTCAGCCACTACACAGCGGTGAGACCGCCCGCTGCGACGTCCTAC 2052
QY      62  LysLeuSerSerProGlnLysAspProThrCysValThrLysGlnPheTrpAlaThr 81
      2053  AAGCTAGCGCGCGCCGACGACGACCGCCCTGACAGACGCGCTTGGGCTAGCAGT 2112
QY      82  LeuAspSerAlaGlyProLeuProAspTYrThrProProGlnLysPheSerPheGlnSer 101
      2113  ATGAGAGCAGCGAGCTGCGCCCGGATTAATGTTCTCTCCAGAGATCTTCCACTTCCACAG 2172
QY      102  ThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGlyLys 121
      2173  CGCTCGATGTGCGGCTGACGATGATCTACAGCGCCGACGCTTGACCGCAGGAG 2232
QY      122  LysTYrProThrValLeuPheIleTYrGlyGlyProGln----- 134
      2233  AAGACCCACCGCTCTCTTGTATATGAGGCGCCCGAGTGACGCTGATTAATCTCC 2292
QY      134  ----- 134
      2293  TTCAAAGCATCAAGTACTTGCGGCTCAACACACTGGGCTCCTCGGAGTACGCGTGTT 2352
QY      134  ----- 134
      2353  GTGATTGACGCGAGGCGCTCTCTCACGCGGCTTCGTTGCAAGGCGCCCTGAAAAAC 2412
QY      135  -----GlyGlnIleGlnLysAspAspGlnValGlnGlyLeuGlnTYrLeuAlaSerArg 152
      2413  CAAATGGCGCAGGTGAGATCGAGACGAGTGGAAGGCGCTGCAAGTTCGAGCGCAGAG 2472
QY      153  TYrAspPheIleAspLeuAspArgValGlyIleHisGlyTYrPheSerTYrGlyGlyTYrLeu 172
      2473  TATGGCTTCATCAGCTGAGCCGAGTTCATGCTGCTGCTGCTTCAACGCGGCGCTTCCTC 2532
QY      173  SerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPro 192
      2533  TCCTCATGGGCTTAATCCCAAGCGCCAGGTGTTCAAGGTGCGCATCGCGGAGTCCCG 2592
QY      193  ValThrLeuTrpIlePheTYrAspThrGlyTYrThrGlnArgTYrMetGlnHisProAsp 212
      2593  GTCACCGTCTGATGGCTTACGACAGGCTACACTGAGCGTACATGAGCGTCCCTGAG 2652
QY      213  GlnAsnGlnGlnGlyTYrTYrLeuGlySerValAlaMetGlnAlaGlyLysPheProSer 232
      2653  AACCAACAGCAGCGCTATGAGGCGGTTCGTGCGCTTCGACGTGAGAGAGCTGCCCAAT 2712
QY      233  GlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnLysValHisPheAlaHis 252
      2713  GAGCCCAACGCGCTTGTATCTCCACGCGCTTCTGACGAAAACTGGCACTTTTCCAC 2772
QY      253  ThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTYrAspLeuGlnIleTYr 272
      2773  ACAAACTTCTCTGCTCCCAACTGATCCGAGCGAGGAAACCTTACCACTCCGATCTAC 2832
QY      273  ProGlnGlnArgHisSerIleArgValProGlnSerGlyGlnHisTYrGlnLeuHisLeu 292
      2833  CCCAACAGAGACAGTATTCCTGCGCGGAGTGGGGGAGACTATGAGATCACGTTG 2892
QY      293  LeuHisTYrLeuGlnGlnLysLeu 300
      2893  CTGCACTTTCTACAGGAATACCTC 2916

```

RESULT 15
 US-10-098-841-100
 ; Sequence 100, Application US/10098841
 ; Publication No. US20020197679A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Xu, Chongjun
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yunging
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Zhang, Jie
/ APPLICANT: Qian, Xiaohong B.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
/ FILE REFERENCE: Polypeptides
/ CURRENT APPLICATION NUMBER: US/10/098,841
/ PRIOR FILING DATE: 2002-03-13
/ PRIOR APPLICATION NUMBER: 09/598,042
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 331
/ SOFTWARE: pc_Fl_genes Version 1.0
/ SEQ ID NO 100
/ LENGTH: 2801
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (158)..(1540)
/ US-10-098-841-100

Alignment Scores:
Pred. No.: 2,71e-127 Length: 2801
Score: 1032.50 Matches: 194
Percent Similarity: 66.208 Conservative: 43
Best Local Similarity: 54.198 Mismatches: 54
Query Match: 61.464 Indels: 67
Gaps: 3

US-10-070-464-3 (1-310) x US-10-098-841-100 (1-2801)
QY      1  PheGlnGlyThrLysAspSerProLeuGlnHisLysLeuTYrValLysTYrValAsn 20
      1752  TTCAGGGGACACAGACAGCGCGCTGAGCACCACTTACGTGCTGATGAGGCG 1811
QY      21  ProGlnValThrArgleuthrhaparglytyrserhissercysalserglnhis 40
      1812  GCGGGGAGATGCTAGCGCTCACACGCGGCTCTCCATAGCTGCTCCATGAGCCAG 1871
QY      41  HisCysAspPhePheIleSerLysTYrSerAsnGlnLysAsnProHisCysValSerLeu 60
      1872  AACTTGAATGTTGCTGACGCCACTACACAGCGGTGAGACCGCCCGCTGCGACGTC 1931
QY      61  TYrLysLeuSerSerProGlnLysAspProThrCysValThrLysGlnPheTrpAlaThr 80
      1932  TCAAGCTAGCGCGCCGACGACGACCGCCCTGACAGACGCGCTTGGGCTAGC 1991
QY      81  IleLeuAspSerAlaGlyProLeuProAspTYrThrProProGlnLysPheSerPheGln 100
      1992  ATGATGAGAGCAGCC-----AAGATCTTCCATTTCCAC 2024
QY      101  SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGlnProGly 120
      2025  ACGCGCTCGATGTGCGGCTTACGCGCATGATCTCAAGCCCGCCTTGACGCGAGG 2084

```

```
QY      121  LysLeuTyrProThrValLeuPhe1LeTyrGlyGlyProGln----- 134
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db      2085  AAGAACACCCACCGCTCTTGTATATGAGGCCCCACGGTGCACTGTGAATTAAC 2144
      |||||:::|||||:::|||||:::|||||:::|||||
QY      134  ----- 134
Db      2145  TCCTTAAGGATCAAGTACTTGCGGCTCAACACTGGCTCTCGGGCTACGCCGTG 2204
      |||||:::|||||:::|||||:::|||||:::|||||
QY      134  ----- 134
Db      2205  GTTGTGATTGACGGGAGGGGCTCTGTCAAGGAGGCTTGCGAAGGGGCTTGAAA 2264
      |||||:::|||||:::|||||:::|||||:::|||||
QY      135  -----GlyGlnIleGlnIleAspAspGlnValGlnIleGlnIleTyrLeuAlaSer 151
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2265  AACCAATGGGCGAGGTGAGATCGAGACCAAGTGGAGGGCTGCGAGTTCGTGGCCGAG 2324
      |||||:::|||||:::|||||:::|||||:::|||||
QY      152  ArgTyrAspPhe1LeuAspLeuAspArgValGlyIleIleGlyTyrPheTyrGlyTyr 171
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2325  AAGTATGGCTTCATCGACTGAGCCGAGTCCATGCTGCTGCTTACGGGGGCTTC 2384
      |||||:::|||||:::|||||:::|||||:::|||||
QY      172  LeuSerLeuMetAlaLeuMetGlnArgSerAsp1LePheArgValAlaIleAlaGlyAla 191
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2385  CTCCTGCTCATGGGGCTAATCCACAAAGCCCAAGGTGTTCAGGTGGCCATCGCGGTGCC 2444
      |||||:::|||||:::|||||:::|||||:::|||||
QY      192  ProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlyArgTyrMetGlyHisPro 211
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2445  CCGGTACCGTCTGTGATGGCTTACGACACAGGGGTACACTGAGCGCTACATGAGACGTCCCT 2504
      |||||:::|||||:::|||||:::|||||:::|||||
QY      212  AspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyIleAspPhePro 231
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2505  GAGAACACACGACGAGCTATGAGCGGAGTTCCTGTGCGCTGTGCACTGGAGAAAGTGGCC 2564
      |||||:::|||||:::|||||:::|||||:::|||||
QY      232  SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPheAla 251
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2565  AATGACCCCAACCGCTTGTGCTTATCTCCACGGCTTCCTGAGCGAAACGTGCACTTTTC 2624
      |||||:::|||||:::|||||:::|||||:::|||||
QY      252  HisThrSer1LeuLeuSerPheLeuValArgAlaGlyIleProTyrAspLeu----- 269
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2625  CACACAACTTCTCTGCTCCCACTGATCCGAGAGGGAAACCTTACCAGCTCCAGGTG 2684
      |||||:::|||||:::|||||:::|||||:::|||||
QY      270  -----GlnIleTyrProGlnGluArgHisSer1LeuArgValPro 282
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2685  GCCCTGCTCTCTCTCTCCCGGACGATCTACCCCAAGAGACACAGTATTCGCTGCCCC 2744
      |||||:::|||||:::|||||:::|||||:::|||||
QY      283  GluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 300
      |||||:::|||||:::|||||:::|||||:::|||||
Db      2745  GAGTGGGGGAGCATATGATGACAGTGTGCACTTCTACAGGAAATACCTC 2798
      |||||:::|||||:::|||||:::|||||:::|||||
```

Search completed: October 16, 2003, 09:08:35
Job time : 262.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 1862.46 seconds
(without alignments)
4045.398 Million cell updates/sec

Title: US-10-070-464-3
Perfect score: 1680
Sequence: 1 FEGRKDPLEHLLVYVSYN.....HLHLHYLQENLGRNALKVI 310

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/CGP2.1/USPFO.spool/US10070464/runat_15102003_113553_24829/app_query.faacta_1.2652
-DB=EST -QFMT=fastcap -SUFFIX=first -MINMATCH=0.1 -FOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464@CGN 1.1 8056 @runat_15102003_113553_24829 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTR -NEG SCORES=0 -WATT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELop=6 -DELext=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssr1:*

29: gb_gssr2:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1605.5	95.6	3143	11 AK016546	AK016546 Mus muscu
2	1605.5	95.6	5517	11 AK029788	AK029788 Mus muscu
3	1309	77.9	1042	12 BM557438	BM557438 AGENCOURT
4	1094.5	65.1	3327	11 AK050021	AK050021 Mus muscu
5	1094.5	65.1	3376	11 AK050023	AK050023 Mus muscu
6	1093.5	65.1	3457	11 AK078301	AK078301 Mus muscu
7	1052.5	62.6	789	10 BG709118	BG709118 602675382
8	1036	61.7	957	13 BG675006	BG675006 AGENCOURT
9	1021.5	60.8	1041	12 B1084090	B1084090 602869453
10	996.5	59.3	746	14 BY751026	BY751026 BY751026
11	991	58.0	910	13 B0675260	B0675260 AGENCOURT
12	962	57.3	788	10 BG163397	BG163397 602338360
13	961.5	57.2	735	13 BU447277	BU447277 603766685
14	939.5	55.9	855	12 B1084885	B1084885 602869453
15	938	55.8	1085	9 AL582306	AL582306 AL582306
16	935.5	55.7	741	14 CB595102	CB595102 AGENCOURT
17	929	55.3	708	13 BU622228	BU622228 UI-H-FH1-
18	929	55.3	735	9 AL043338	AL043338 DKFZP4340
19	929	55.3	910	10 BG479035	BG479035 602526111
20	929	55.3	1002	10 BE733691	BE733691 601569288
21	929	55.3	1201	9 AL542617	AL542617 AL542617
22	925	55.1	776	14 CB522171	CB522171 UI-M-GH0-
23	925	55.1	800	12 BG921267	BG921267 602824107
24	925	55.1	905	14 CB209486	CB209486 AGENCOURT
25	918	54.6	746	14 BY764370	BY764370 BY764370
26	902.5	53.7	695	13 BG715866	BG715866 602675282
27	887.5	52.8	968	13 B0671635	B0671635 AGENCOURT
28	882	52.5	804	13 B0687687	B0687687 UI-CF-EC1
29	877	52.2	542	13 BX259029	BX259029 BX259029
30	864	51.4	617	10 BE789604	BE789604 601481552
31	862	51.3	674	13 BU229303	BU229303 603800110
32	855	50.9	603	12 B1455842	B1455842 603171060
33	845.5	50.3	1082	13 BU239476	BU239476 603322338
34	845	50.3	493	9 AA278625	AA278625 z878911.r
35	838.5	49.9	674	10 BE888665	BE888665 601513061
36	829	49.3	484	9 AA417787	AA417787 zV01a01.r
37	822	48.9	797	14 CA749361	CA749361 UI-M-FY0-
38	809.5	48.2	639	10 BG165205	BG165205 602344074
39	801	47.7	852	13 BU236501	BU236501 603408785
40	796	47.4	760	9 A1917735	A1917735 t11id07.x
41	792	47.1	627	10 BG752328	BG752328 602730802
42	766.5	45.6	753	9 AL040398	AL040398 DKFZP434A
43	762.5	45.4	853	12 B123892	B123892 602941035
44	760	45.2	897	13 BU191638	BU191638 AGENCOURT
45	738.5	44.0	658	10 BE388695	BE388695 601286345

ALIGNMENTS

RESULT 1
LOCUS AK016546 3143 bp mRNA HTc 05-DEC-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog
[Homo sapiens], full insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS
SOURCE HTC, CAP trapper.
ORGANISM Mus musculus
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

99279253

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

PUBMED

20499374

11042159

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsumai, T., Teshiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nieline, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M., Yoneda, Y., Isikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawaj, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kirz, A. and Hayashizaki, Y.

Riken integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

MEDLINE

PUBMED

20530913

11076861

4

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Atkawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gogobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flischiemann, N., Gaasterland, T., Gissi, C., King, B., Kochiya, H., Kusbi, P., Lewis, S., Marcou, Y., Nikaido, I., Peeble, G., Quackenbush, J., Schriml, L.M., Stabill, F., Suzuki, R., Tomita, M., Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boiffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bul, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotchman, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Macchiommi, L., Mashima, J., Mazzarelli, J., Mombers, P., Noridone, P., Ring, B., Rinnwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Sey, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-Oka, K., Wang, K.K.H., Weitz, C., Whitlaker, C., Williams, J., Wyshna-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsuki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection

Nature 409 (6821), 685-690 (2001)

JOURNAL

MEDLINE

PUBMED

21085660

11217851

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3143)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hisatsue, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Osa, C., Saito, H., Saito, R., Sasaki, C., Sasaki, R., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogahe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Yamatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATTCGAGTCAAGACCTCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGATTCGAGTCAAGTAAATTAATTCACCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified plasmidscript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source

location/Qualifiers

1..3143

organism="Mus musculus"

mol_type="mRNA"

strain="C57BL/6J"

db_xref="FANTOM_DB:4932434F09"

db_xref="MGJ:1907554"

db_xref="taxon:10090"

clone="4932434F09"

sex="male"

tissue_type="testis"

clone_lib="RIKEN full-length enriched mouse cDNA library"

dev_stage="adult"

327..3005

note="unamed protein product; DIPEPTIDYL PEPTIDASE 8 homolog (Homo sapiens) (SPTR|Q9HBM5, evidence: FASTY, 95.7%id, 100%length, match=2649)

putative"

codon_start=1

protein_id="BAB30295.2"

db_xref="GI:26382128"

db_xref="MGJ:1921638"

translation="MKIPSRRCNMAAMAEIQLGVEIFETAECEGNGESQDRKLEP FYEYRISWSQKLLADTRKRYHGYYAKAKAHDFMFVKRTPDRHSORVYLLANGSEN RAYTLFYSEIPIKTINRAAYLMLSWKPLDIFQATLDGWSREBELREKRICTGVI AADVHGSGEPLFOAGSGIYHIKDGGPHGFQTPNPINLVETSCPIIMDPKICPADP DMIAFIHSDINISNLVTRREERITVYHNLAMEBDEPSAGVATVLDEEFYRSGYK YMMQPOAKERTPSGKKILRIILEENDESEVRIHYTSPMLSTRADSRYPKGTANRK VTRMSRIYVDAAGCITDIVDKELVQFEILFEGVEETIAAGCTPEKAMSLILDS QTHQIVLISPELPIVEDDAMDRQRLIESVPSTVPLIIYEETDWINIHDFHVF PQRHEDIEIFIFASECKTGRHLKYKTLISKEKYSKSGSLPAPSDFOCKPIKEITIT TSGVEWELGRHGINSINWDEARKLVFEGTGSQPLEHLVYTSANPEVVRLLDGRYS HSCCLSHWCFPLSKYSNOKRPHCVSLKSSBEDPDVHTKEWATILDSAGLDPDY TPTLPIFEFESTGTCTGMLYKPHDLOPGKYPVLPYIYGPOVLNNPFQKGYRFR LNTLASIGYVVVVINDRNGSCHRLKPEGAKRYMMGOIEIDPOVEGLQYLAISQVDIDL DRGICRMSYCGYLSLMLALMQRSDIFRYALAGAPYTIMITYDTGTYTRWGHDPONCO GRTYGVVAMQAEKPESEPNLLHGLGUDSEVFAHRTSILSPFVLRGKPYDQITPQ ERHSIRVSGEYGEVLEHLHYLDENLGSRTAAALKVT"

BASE COUNT

894 a 688 c 773 g 788 t

ORIGIN

Alignment Scores:

Pred. No.: 1.63e-169 Length: 3143

Score: 1605.50 Matches: 302

Percent Similarity: 84.96% Conservative: 3

Best Local Similarity: 84.12% Mismatch: 5

Query Match: 95.57% Indels: 49

DB: 11 Gaps: 1

US-10-070-464-3 (1-310) x AK016546 (1-3143)

Db |TTTGAAGGACCAAGACTCTCTTGGAGCATCACTTAGTGACCGAGTTATGCAAC 1985
 1926
 Oy |||||ProGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
 |||||1986 CCGAGGAAGTGTGGCTGAGTGCCTGGCTACTCACTCTGGCTGCACCGG 2045
 |||||41 HisCysAspPhePheIleSerIleTyrSerAsnGlnIleAsnProHisCysValSerLeu 60
 |||||2046 CATTGGAATTCTTCTCAATAGTAAGTACACCMACCGAAGAAATCCACACTGTGTCTTC 2105
 |||||61 TyrIleLeuSerSerProGluAspAspProThrCysValSerThrGlnPheTrpAlaThr 80
 |||||2106 TACAAACTCTCAAGTCTGAGATGAGCCAGATTCAATAAACAAGAAATTTTGGCCACC 2165
 |||||81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 100
 |||||2166 ATTTGGATTACAGCAGGCTCTCTCTGCTACACCCCTCCAGAAATTTTCTTTTGA 2225
 |||||Oy 101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLeuProHisAspLeuGlnProGly 120
 |||||2226 ACTTACTAGTGAATTTACACTGTATGGAAATGTTGTATTAAGCTTCATGACTTACAACTGGA 2285
 |||||Oy 121 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
 |||||2286 AAGAAATACCCCATGTGTATTATCATATATGTGTGTCCCGAGTGCAGCTGTGAACAT 2345
 |||||Oy 134 ----- 134
 |||||Db 2346 CGGTTTAAAGAGTCAAGTATTTCCGCCCTGAACACCCCTGCGCTCCCTGGTTATGTGTT 2405
 |||||Oy 134 ----- 134
 |||||Db 2406 GTGGGTATGACAAACAGGGGATCTGTCAACGAGACTTAATTTGAAGCCCTTTAA 2465
 |||||Oy 135 -----GlyGlnIleGluIleAspAspGlnValGlyIleGlnTyrLeuAlaSer 151
 |||||Db 2466 TATAAATGTGGTCAATATGAATATCGATGATCAAGTGAAGGACTCCAGTACCTAGCATCT 2525
 |||||Oy 152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyTyr 171
 |||||Db 2526 CAGTATGACTTCAATGACTTGGATTCAGAGGGGATCCAGGGCTGCTATGGTGCATC 2585
 |||||Oy 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
 |||||Db 2586 CTCTCCCTGATGGCATTAATGACAGAGTGGATATCTCCGGTGGCTATTCGTGGCGCC 2645
 |||||Oy 192 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
 |||||Db 2646 CCAGTCAACCTGTGATCTTCTATGATACAGATACACGAGCGCTATATGGGTCACTCT 2705
 |||||Oy 212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPhePro 231
 |||||Db 2706 GACCGAAGTGAACAGGGCTACTACTTAAGATCTGTGGCCATGCAACGGAGAAATTTCC 2765
 |||||Oy 232 SerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluLysAsnValHisPheAla 251
 |||||Db 2766 TCAGAACCAACCGGTATCTCTCTTGGATGGGTCTTGGATGAGAAATTTCACTTTGCA 2825
 |||||Oy 252 HisThrSerIleLeuLeuSerPheLeuValAlaGlyLysProTyrAspLeuGlnIle 271
 |||||Db 2826 CACACACAGTATATGCTGAGTGTATTTAGTGAAGGCGTGAACCAATATACCTTACGATC 2885
 |||||Oy 272 TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis 291
 |||||Db 2886 TATCTTCAGAGAGGACACGATCAGAGTTCTCTGACTTGGAGAAACACTATGAACGCGAC 2945
 |||||Oy 292 LeuLeuHisTyrLeuGlnGluLysAsnLeuGlySerArgIleAlaAlaLeuLysValIle 310
 |||||Db 2946 CTGGCTCACTACTCTTCAAGAGAACTTGATCGCGATCGCTCTGAAAGTGAATA 3002

RESULT 2
 AK029788

LOCUS AK029788 5517 bp mRNA linear HTC 05-DEC-2002
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:493056OC15 product:DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens], full insert sequence.
 ACCESSION AK029788
 VERSION AK029788.1 GI:26081520
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 Carninci, P., and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20493774
 PUBMED 11042159
 REFERENCE 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Teshiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, B., Matsubuchi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL Genomics 34, 175-184 (2000)
 MEDLINE 1076861
 PUBMED 11076861
 REFERENCE 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Maceno, Y., Nikaudo, I., Resole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldairelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Meshima, J., Mazzarelli, J., Nombela, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitteker, C., Wilming, L., Wyshaw-Borja, A., Yoshida, K., Haegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851
 REFERENCE 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 12042159
 PUBMED 12042159
 REFERENCE 6 Adachi, J., Aizawa, K., Akimura, T., Aizawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hayashizaki, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirokawa, T., Hironaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-fes@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers

FEATURES
source
1..5517
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4930560C15"
/db_xref="taxon:10090"
/clone="4930560C15"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
2221..4837
/note="DIPPTIDYL PEPTIDASE 8 homolog (Homo sapiens) (SPR) (OHBM5, evidence: FAST, 95.7%id, 100%length, match=2649)
putative"

BASE COUNT 1514 a 1224 c 1230 g 1549 t
ORIGIN
Alignment Scores:
Pred. No.: 3,91e-169 Length: 5517
Score: 1605.50 Matches: 302
Percent Similarity: 84.96% Conservative: 3
Best Local Similarity: 84.12% Mismatches: 5
Query Match: 95.57% Indels: 49
DB: 11 Gaps: 1
US-10-070-464-3 (1-310) x AK029788 (1-5517)

QY 1 PheGluGlyThrTyAspSerProLeuGluHisLeuTyValValSerTyValAsn 20
DB 3758 TTGAAAGGACCAAAACCTCTCTTGGAGCATCAGCTGAGTGACCGATTATGCAAC 3817
QY 21 ProGluGluValThrArgLeuThrAspArgGlyTySerHisSerCyAsiSerGin 40
DB 3818 CCGAGGAAGTGGTGGCTGAGCTGACGCTGCTACTCAACATCCGCTGCTCAGCCG 3877
QY 41 HisCyAspPhePheLeuSerTySerAsnGluHisAsnProHisCyValSerLeu 60
DB 3878 CATTGAGCTTCTTCTAAGTAAGTACACCAACCAAAATCCACAGTGTGCTCC 3937
QY 61 TyrTyLeuSerSerProGluAspPheProThrCyValTyThrTyGluGluPheTrpAlaThr 80
DB 3938 TCAAAACTCTCAAGTCTGAGATGACCCAGTTCATAAACAAAGAAATTTGGGCGACC 3997
QY 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProGluGluPheSerPheGlu 100
DB 3998 ATTGGATTACGAGGCTCTCTCTGACTACACCCCTCCAGAAATTTCTTTTGA 4057

QY 101 SerThrThrGlyPheThrLeuTyGlyMetLeuTyTyAspProHisAspLeuGlnProGly 120
DB 4058 ACTACTACTGAGATTACACTGATGAAATGTTGTAAGCCCTCATGACTACCACTGGA 4117
QY 121 TyLeuTyProThrValLeuPheLeuTyGlyGlyProGln----- 134
DB 4118 AAGAAATACCCCACTGTATTATATATATGATGATGCCAGGTGACGCTGTGAACAT 4177
QY 134 ----- 134
DB 4178 CCGTTAAAGAGTCAAGTATTCGCCCTGAACACCTGGCTCCTGGATTATGTGTT 4237
QY 134 ----- 134
DB 4238 GTGTGATATGACACAGGGATCTGTGACGAGACTTAATTGAAGCGCCTTTAA 4297
QY 135 -----GlyGlnIleGluIleAspArgIleValGluGlyLeuGlnTyLeuAlaSer 151
DB 4298 TATTAATGCGTCAATATGAAATGATGATCAAGTGAAGGACTCCAGTACCTGATCT 4357
QY 152 ArgTyAspPheIleAspLeuAspArgValGlyIleHisGlyTyrSerTyGlyTyTy 171
DB 4358 CAGTATGACTTCAATGACTGATGATGATGAGGAGATCCAGGCTGCTATGATGCTAC 4417
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
DB 4418 CTCTCCCTGATGAGCTTAATGACAGAGTGCATCTCCGGGTGCTATGCTGGGCGCC 4477
QY 192 ProValThrLeuThrIlePheTyThrPheThrGlyTyThrGluArgTyMetGlyHisPro 211
DB 4478 CCAAGTACCCCTGTGATCTCTATATACAGATACACGAGCGCTATATGGGTACCTC 4537
QY 212 AspGlnAsnGluGlnGlyTyTyTyLeuGlySerValAlaMetGlnAlaGluTyPhePro 231
DB 4538 GACCAAGATGACAGAGGCTACTACTTAGATCTGTGGCCATCTCAAGCGAGAGTCCCC 4597
QY 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAla 251
DB 4598 TCAGAACCAACCGGTTACTCTCTGACGAGGCTTGGAGAGAAATTTCACTTGCA 4657
QY 252 HisThrSerIleLeuLeuSerPheLeuValAspAlaGlyTyProTyThrAspLeuGlnIle 271
DB 4658 CACACCAAGTATATCTGATGATTTTATGAGAGGCTGGAACCCATATGACTTACGATC 4717
QY 272 TyrProGluGluArgHisSerIleArgValProGluSerGlyGlnHisTyGlyLeuHis 291
DB 4718 TATCTTCAGGAGAGGACACGATCAGATCTCTGATCTGAGAACACATGATGACGAC 4777
QY 292 LeuLeuHisTyTyLeuGlnGluAsnLeuGlySerArgIleAlaIleAlaLeuValIle 310
DB 4778 CTGCTCCACTACTTACAGAGAACTTGATCGCTATGCTGCTGCTGAAAGTATGA 4834
RESULT 3
BMS57438
LOCUS BMS57438 1042 bp mRNA linear EST 20-FEB-2002
DEFINITION AGCCOVRT 6578992 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466759
ACCESSION BMS57438
VERSION BMS57438.1 GI:18799430
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1042)
NIH-MGC http://mgi.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
UNPUBLISHED
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://www.llnl.gov/imaging>

BASE COUNT	279 a	223 c	234 g	303 t	3 others
ORIGIN					

Alignment Scores:	
Pred. No.:	8.51e-137
Score:	1309.00
Percent Similarity:	78.96%
Best Local Similarity:	78.05%
Query Match:	77.92%
DB:	12
	Gaps: 2
	Length: 1042
	Matches: 256
	Conservative: 3
	Mismatches: 14
	Indels: 55

US-10-070-464-3 (1-310) X BM557438 (1-1042)

[illegible]

Oy		195	-----G VgInIIEguIlIeAsPaSPglvAlGlIgUgYLeugInYrIleuaIAseR	151
Db		601	TATATAAATGGGTCAAATTGAATTAAGCATCAGGTGAAGAAGCACTTCAATATCTTAGCTTCT	660
Oy		152	ArgTYraAPPeIIlaEPleuaASpaRvaIGlyIleHieglYTPrSerTYrgLYgTYr	171
Db		661	CGATRTAGATTTTCATTGACTTATGATCGTGCGGACCTCACGGCTGGTCTTATGAGAGATAC	720
Oy		172	LeuSerIeuMeCaIaleMeuEGlnarSeraSpIIlePheargVaIAIlealagIYAIA	191
Db		721	CTCTCCCTGATGGCATTAAATGCAGAGGTACGATATCTTTCAGGCGTGTGATTGCTGGGGCC	780
Oy		192	ProValThrlEuTrPIIlePheTYraSPThrGIyTYrThrgIuargTYrmeGIYHISPro	211
Db		791	CCAGTCATCTGTGTGAATCTTCTTAGAATACAGAGATACCGGTACTTATATGGGTCAACCCT	840
Oy		212	AapGINaengIugIngIYrTYrTYrLeUGlYSerValaIameGINaIaGIuIVasPhaePro	231
Db		841	GACCCGAATGACAGGGCTATTACTTAMGATCTGTGGCCATGCAGACGAAAAAGTCCCC	900
Oy		232	SergIuProaenArGLeuLeuLeuLeuHieglY-PhelEuaSPgluAnval-HIaPhea	251
Db		901	TCTGACCAATATGCATTACTGCTTTTACATAGCTTTCTCGATAGAAATGCCATTGTTGG	960
Oy		251	IahSerHser-IleLeuLeuSer-PheLeuValaIArgIaGlyLysProTYraPleu---	269
Db		961	CACATACCACGAGATATTACTGAGTTTTTTNTNGAGAGGGCTGAAAAGCCCATATGAT	1020
Oy		270	-----GlnIeTYrPro 273	
Db		1021	TTTACAGAACTATCTTCT 1038	
RESULT 4				
AKO50021				
LOCUS				
DEFINITION				
AKO50021				
Mus musculus adult male liver tumor cDNA, RIKEN full-length				
enriched library, clone:CT30003012 product:DipePTIDYL PEPTIDASE 9				
homolog (Homo sapiens), full insert sequence.				
OA050021				
AKO50021.1 GI:26340743				
HNC; CAP trapper.				
Mus musculus (house mouse)				
SOURCE				
ORGANISM				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE				
AUTHORS				
1 Carninci,P. and Hayashizaki,Y.				
High-efficiency full-length cDNA cloning				
Meth. Enzymol. 303, 19-44 (1999)				
JOURNAL MEDLINE				
PUBMED				
10349636				
REFERENCE				
AUTHORS				
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,				
Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
Normalization and subtraction of cap-trapper-selected cDNAs to				
prepare full-length cDNA libraries for rapid discovery of new genes				
Genome Res. 10 (10), 1617-1630 (2000)				
JOURNAL MEDLINE				
PUBMED				
20499374				
11042159				
REFERENCE				
AUTHORS				
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,				
Komno,H., Akiyama,J., Nishi,K., Katsunai,T., Tashiro,H., Itoh,M.,				
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishise,T., Harada,A.,				
Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kaishiwagi,K.,				
Fujiwake,S., Inoue,K., Togawa,Y., Tanaka,M., Ohara,E., Wataniki,M.,				
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,				
Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.				
RIKEN integrated sequence analysis (RISA) system-384-format				
sequencing pipeline with 384 multicapillary sequencer				
Genome Res. 10 (11), 1757-1771 (2000)				
JOURNAL MEDLINE				
PUBMED				
20503913				
11076861				
REFERENCE				
4				

AUTHORS
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kado, S., Yamane, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Boujunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohseuki, S. and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection

JOURNAL
Nature 409 (6821), 685-690 (2001)

PUBMED
21085660

REFERENCE
11217851

AUTHORS
5

TITLE
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

JOURNAL
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

REFERENCE
Nature 420, 563-573 (2002)

AUTHORS
6 (bases 1 to 3327)

TITLE
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE
Direct Submission

JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES
Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.

source
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Location/Qualifiers
1. 3327
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:C730003D12"
/db_xref="taxon:10090"
/clone="C730003D12"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"

CDS
/clone_11b="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
176..2764
/note="unnamed protein product: DIPEPTIDYL PEPTIDASE 9 homolog [Homo sapiens] (SPTRLA147179, evidence: PASTY, 92410, 99.61length, match=2580)
putative"
/codon_start=1
/protein_id="BAC34034.1"
/db_xref="GI:26340744"
/translation="MCSGVSPVEOVAAGMDPTAARFCYOKXSMQIESTGRKSSGLVSKAPRHPFVQKPDGSGPHSRRLTYLGMFYGRSSLSLYSELPKVRKREALLSKQMLIDHFOATPHNGVYSRESELRERRLDGPDPFSESGFLFQASNSLFLCDDGKGMFVSPMPKPLEIKTDCSFRMPKCPADPAFSEFINSNDLVANILEEERBLTFCQOSACVLNDPKSAGVATVIOEFPFGCMCFPASMSESGLTKTRILYBEVDSSEVIVHPSPALREKRTDSYRPRGSKAPKIALKLABIOTDQKIVSSCEKLVQPFSSLPFKVETIARAGCTRGKTKYMANFLRPOORLDVLLPRLFPAYASEARQAPARVPKRVQRPVATYEVTVNWIWHDIFRFPQACQDQCFILABEKTFGCHLYVTVELKTRKDYDTEPLSPTEDEFKCPKEKVALTSGMEVLSRHSKTIWNEOTKLIVFOGTRKDPRLSHHLVYVSYESAGEIVRLTLTGFSHSCMSQSFDMFVSHSVSTPPCVHVVYKLSGDDPDLKQPRFVASMMEANCPDPVPEIPIFHTRADVQYGMITYKPHILOPGRKHPVLFVYGGPOVLNNSPFKIGYLRNLTLASIGVAVVVIDGRSCOGKAREGALKNQMGQVEIEDVEGLQVYAEKYGKIDLSRVAIHGMSYCGPLSIMGLIHQVFKVATIAAPVATVMAVDGTERTMDYPERNNQSGYEGVALHAEKLPENRNLILHGFGLDENVHFHTNPLVLSQLIRAGKPYOLOIYPNRHSIRRESEGEHVEVTLH FLOEHL"

BASE COUNT
702 a 1015 c 919 g 691 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	3327
Score:	8,33e-112	3327
Percent Similarity:	1094.50	200
Beet Local Similarity:	69.348	Conservative: 42
Query Match:	65.154	Mismatches: 58
DB:	11	Indels: 49
		Gaps: 1

US-10-070-464-3 (1-310) x AK050021 (1-3327)

QY 1 PheGluGlyThrLysAspSerProLeuGluHisLysLeuValValSerTyValAsn 20
1715 TTTCAAGGTACAAAGACACACCGCTGGAACATCACTTATGTGTACGTACAGTCA 1774

QY 21 ProGluGluValThrArgLeuThrAspArgGlyTySerHisSerCyAsylSerGln 40
1775 GCAGGGAGATGTGGCGGCTCACACGCTCGGCTTCCACAGCTGCTCCAGACAG 1834

QY 41 HisCyAspPhePheIleSerIyTySerAenglnLysAsnProHisCyValSerLeu 60
1835 AACTTCGACATGTCGTGAGTCACTACACAGAGTGCAGCACCCTGTGTACATG 1894

QY 61 TyrlsLysSerSerProGluAspAspProThrCyLysLeuThrLysGluPheTpaIaThr 80
1896 TACAACTAGAGGGCCCGCATATATACCACTGCACACAAACACAGCTTCTGGCCAC 1954

QY 81 IleLeuAspSerAlaGlyProLeuProAspTyThrProProGluIlePheSerPheGln 100
1955 AGATGAGAGGACGCCAATGCCCCCAGACTATGTGCCCCCTTGAGATCTTCACCTTCCAC 2014

QY 101 SerThrThGlyPheThrLeuTyrlYwctLeuTyrlLysPProHisAspLeuGlnProGly 120
2015 ACCCGTCAGAGCTGACGCTTACGGCATGACCTTCAMACCCACACACCTGCAACCTGGG 2074

QY 121 LysLysTyrlProThrValLeuPheIleTyrlGlyGlyProGln 134
2075 AGGAAGACACCCACTGTGTCTTTGTCTATGGGGGCCACACAGCTGACGTGCTGAACAC 2134

QY 134 2135 TCCTTTAAGGAGCATAAATACCTGCGGCTAAATACACTGACATGCTTGGGCTATGCTGG 2194
134 134

Db	2135	GTGGGATGATCGATGTCGGGGGCTCTCGTCAAGCGGGGCTTGACCTTGAAGGGGGCCCTGA	225
Qy	135	-----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer	151
Db	2255	AATCAAAATGGGCGAGGTAGAGATTAGAGACCAAGGTGGAAGGCTTGACATACCTGGCTTAG	2314
Qy	152	ArgTyrAspPheIleAspLeuAspArgValGlyIleHisgLyTyrSerTyrGlyGlyTyr	171
Db	2315	AAGTATGGCTTCATTGACTTGAAGCCGAGTCCGACATTCATGGCTGGTCTTACGCGGGCTTC	2374
Qy	172	LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla	191
Db	2375	CTCTACATCATGGGCTCATTCACAAGCAACAAGTTCGAAGTATGCCATTTGGGGCCCT	2434
Qy	192	ProValIleThrLeuTyrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro	211
Db	2435	CTGTGCACGTGTGATGATGACCTTCATGACACAGGGTACACAGGAACGATACATGATGATGCC	2494
Qy	212	AspGlnAsnGluGlnGlyTyrTyrTyrLeuGlySerValAlaMetGlnAlaGluIlyPhePro	231
Db	2495	CAAAATTAACACGACCAAGGCTATAGGACAGGGTCTGTAGCCCTGATGTGAGAAAGCTGCC	2554
Qy	232	SerGluProAsnArgLeuLeuLeuLeuHisgLyPheLeuAspGluAsnValHisPheAla	251
Db	2555	AATGAGCCTAACCCGCTGCTATTACTCCACAGGCTTCCTGACAGAAAGTTCATCTCTTC	2614
Qy	252	HisThrSerIleLeuLeuSerPheLeuValArgAlaGlyIlyProTyrAspLeuGlnIle	271
Db	2615	CACACAAATTTCTCGTGTCTCCACGTGATCCGACGACGAAAGCCATACACACTTCAGATGC	2674
Qy	272	TyrProGlnGluArgHisSerIleArgValProGluSerGlyGlnHisTyrGluLeuHis	291
Db	2675	TACCCAAACGAGACCATGACATCCGCTGCGCCGCGAGTCCGAGACCATTAGAGGTGACG	2734
Qy	292	LeuLeuHisTyrLeuGlnGluAsnIleu	300
Db	2735	CTGCTGCACTTTCTGACGAAACACCTG	2761
RESULT 5			
AK050023			
LOCUS			
DEFINITION	AK050023	3376 bp mRNA linear HTC 05-DEC-2002	
		Mus musculus adult male liver tumor cDNA, RIKEN full-length	
		enriched library, clone:G73003H19 product:DIPEPTIDYL PEPTIDASE 9	
ACCESSION	AK050023	homolog (Homo sapiens), full insert sequence.	
VERSION	AK050023.1	GI:26093756	
KEYWORDS		HTC; CAP trapper.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE		Bukacina, P., Itoh, M., Mizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE		1 Carninci, P. and Hayashizaki, Y.	
TITLE		High-efficiency full-length cDNA cloning	
JOURNAL		Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE		99279253	
PUBMED		10349636	
REFERENCE		2	
AUTHORS		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
		Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
		Normalization and subtraction of cap-trapper-selected cDNAs to	
		prepare full-length cDNA libraries for rapid discovery of new genes	
TITLE		Genome Res. 10 (10), 1617-1630 (2000)	
JOURNAL		20499374	
MEDLINE		11042159	
PUBMED		11042159	
REFERENCE		3	
AUTHORS		Shibata, K., Itoh, M., Mizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,	
		Konno, H., Akiyama, U., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,	
		Sunji, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,	
		Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,	
		Fujiwaka, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Matsubiki, M.,	
		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,	
		Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.	

TITLE
JOURNAL
MEDLINS
PUBMED
REFERENCE
AUTHORS

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

TITLE
JOURNAL
MEDLINS
PUBMED
REFERENCE
AUTHORS

4
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Atakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aikawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamane, I.,
 Saito, T., Okazaki, Y., Gojodori, T., Bono, H., Kaubukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Flischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiya, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, P., Baresch, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., But, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D.A., Kamita, M., Lee, N.H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzaletti, J., Mommaerts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
 Toyokawa, K., Wang, K.H., Weitz, C., Whiteaker, C., Wilming, L.,
 Wrayshew-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S.
 and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 02108560
 11217851

TITLE
JOURNAL
MEDLINS
PUBMED
REFERENCE
AUTHORS

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3376)
 Adachi, J., Aizawa, K., Akimura, T., Atakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hixomoto, K., Hirooka, T., Hirozane, T.,
 Hori, F., Imocani, K., Iehi, Y., Itoh, M., Kagawa, I., Kaubukawa, T.,
 Kachi, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, K., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Yumatsun, M., and Hayashizaki, Y.
 Direct Submission
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 Riken Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
 URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissue was provided by William A. Held, Roswell Park Cancer
 Institute, Department of Molecular and Cellular Biology, Elm and
 Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.
 URL: http://genome-gsc.riken.go.jp/
 URL: http://fantom-gsc.riken.go.jp/
 Location/Qualifiers

FEATURES
SOURCE

1. 3376
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"

```

/db_xref="PANTOM,DB:C730003H19"
/db_xref="taxon:10090"
/clone="C730003H19"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
235_2824
/note="DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens)
(SPTRI/AA147179, evidence: FASTA, 92%id, 99.6%length,
match=2580)
putative"
misc_feature

BASE COUNT      698 a      1046 c      932 g      700 t

ORIGIN

Alignment Scores:
Pred. No.:      8,53e-112      Length:      3376
Score:          1094.50      Matches:      200
Percent Similarity: 69.34%      Conservative: 42
Best Local Similarity: 57.31%      Mismatches: 58
Query Match:      65.15%      Indels:      49
DB:              11      Gaps:      1

US-10-070-464-3 (1-310) x AK050023 (1-3376)

QY      1 PheGluGluThyThrAspSerProLeuGluHisLeuTyValSerTyValAsn 20
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      1775 TTTCAGAGTACAAAGACACACCGCTGAAACATCCTTATGTGTGCTACCTCA 1834
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      21 ProGluValuAlaThrArgLeuThrAspArgGlyTyrSerHisSerCysGlySerGln 40
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      1835 GCAGGGGAGATCGTGGGGCTCACACGCTCGGGCTTCCACAGCTCCATGAGCGAC 1894
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      41 HisCysAspPhePheHisSerLysSerAsnGlnLysAsnProHisCysValSerLeu 60
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      1895 AGCTTCGACATGTTCTTGAGTCACTACAGAGTGTGACAGCGCACCTTGATCATGTG 1954
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      61 TyrLeuLeuSerSerProGluAspAspProThrCysLeuThrArgGluPheTPalaThr 80
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      1955 TACAACTGAGCGGCCCGCATGATGACCACTGACAGAACACACCGCTTCTGGCGCAGC 2014
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      81 IleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGlu 100
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2015 ATGATGAGCGCACCAATTGCCCCAGACATGCTGCCCTGAGATCTTCCACTTCCAC 2074
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      101 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLeuProHisAspLeuGlnProGly 120
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2075 ACCCGTGCAGACGTGCAGCTCTACGGCATATCTACAAAGCAACACCTGCAACCTGGG 2134
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      121 LysLeuTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2135 AGGAAGACACCCCACTGTGCTTTGTCTATGGGGGCCCAAGGTCAGTGTGGAACAAC 2194
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      134 ----- 134
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2195 TCCTTAAGGGCATCAATACTGCGGCTAAATACACTGGCATCTTGGCTATGCTGTG 2254
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      134 ----- 134
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2255 GTGGTATCGATGCTCGGGCTCCTGTGCAGCGGGGCTGCACCTTCGAGGGGGCCCTGAAA 2314
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 151
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2315 AATCAATATGGCGCAGGTGAGATTTAGAGACCAAGTGAAGCTTGCGATCGTGGCTGAG 2374
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      152 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPheSerTyrGlyGlyTyr 171
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2375 AAGTATGCTTCACTGATTCAGCCGAGTGCATTCATGCTGCTTCAAGCGGCTTC 2434
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 191
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2435 CTCTCACTCTATGGGGCTCATCAACAACCAAGTGTTCAGGTAGCATTTGGGGGCT 2494
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

QY      192 ProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro 211
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2495 CCTGTACTGTGTGATGAGCGCTATGACAGAGGTACAGGAAGATACATGATGTCGCC 2554
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      212 AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhePro 231
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2555 GAAATATACAGCAGCAAGCATATGAGCGAGTCTGTGATCCCTGATGTGGAAGCTGCC 2614
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      232 SerGluProAspArgLeuLeuLeuLeuGlyPheLeuAspGluAsnValHisPheAla 251
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2615 AATGAGCTTACCGCTGCTTATCTCCACGCTTCTGGACGAAACGTTCACTTCTTC 2674
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      252 HisThrSerIleLeuLeuSerPheLeuValAlaArgAlaGlyLysProTyrAspLeuGlnIle 271
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2675 CACACAAATTTCTGTGTCCTCCAGCTGATCCGAGCAAGGACATACCACTTCAGATC 2734
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      272 TyrProGluGluArgHisSerIleArgValProGluSerGlyGluHisTyrGlyLeuHis 291
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2735 TACCAAGAGAGAGACATGATGATCCCTGCCGAGTCCGAGAGCATTAAGAGTGCACG 2794
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY      292 LeuLeuHisTyrLeuGlnGluAsnLeu 300
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB      2795 CTGCTCACTTCTTGCAAGAACCTTG 2821
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 6
AK078301
LOCUS
DEFINITION
MUS musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430584g1 product:DIPEPTIDYL PEPTIDASE 9
homolog [Homo sapiens], full insert sequence.
AK078301
VERSION
AK078301.1 GI:26347124
KEYWORDS
HTC; CAP trapper.
MUS musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Carninci,P., Shibata,Y., Hayatsu,N., Sugihara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
20499374
PUBMED
11042159
REFERENCE
AUTHORS
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Katsunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsubara,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
MEDLINE
20530913
PUBMED
11076861
REFERENCE
AUTHORS
Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Aizawa,T., Hara,A., Fukunishi,Y., Kono,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamana,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadoya,K., Matsuda,H., Ashburner,M., Balasov,S., Casavant,T.,
Fleischmann,M., Gasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G.,

```

Quackenbush, J., Schriml, L.M., Staab, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarrelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wyszynski, B., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection

JOURNAL
MEDLINE
21085660
11217851

REFERENCE**TITLE**

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
REFERENCE
NATURE 420, 563-573 (2002)
6 (bases 1 to 3457)

Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, D., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasakawa, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaki, A., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE

Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

CDS

1..3457
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:6430584G11"
/db_xref="taxon:10090"
/clone="6430584G11"
/sex="male"
/tissue_type="Olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
182..2770
/note="unnamed protein product, DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens) (SPTR) AL47179, evidence: FASTY, 92%ID, 99.6%length, match=2580)
putative"
/codon_start=1
/protein_id="BAC37211.1"
/db_xref="GI:26347125"
/translation="MCSGVSPVEQVAAQMDMTAARFCVQKHSMDGLSIHGRKSS GLVSKAPHDFOFVOKPDSGPHSHRLYYLGMVPSRENSLLSYIPKKVRKALLLL

BASE COUNT
ORIGIN 766 a 1033 c 926 g 732 t

Alignment Scores:
Pred. No.: 1,15e-111 Length: 3457
Score: 1093.50 Matches: 200
Percent Similarity: 69.34% Conservative: 42
Best Local Similarity: 57.31% Mismatches: 58
Query Match: 65.09% Indels: 49
DB: 11 Gaps: 1

US-10-070-464-3 (1-310) x AK078301 (1-3457)

QY 1 PheGUGUlyThLyaApSerProLeuGluHSHleuTYrValValSerTYrValaIn 20
DB 1721 TTTCAAGGATCAAGACACACCGCTGGAACATCACTTATGTGTGACGATCAAGTCA 1780
QY 21 ProGUGUlyValThArgleuThraPapaArgGlyTYrSerHisSerCyAsIleSerGin 40
DB 1781 GCAGGAGATGCTGGCTGCTACACCGCTGCTTCCACAGCTGCTCCATGACCCAG 1840
QY 41 HisCyAsPhePheHisSerIleSerIleSerIleSerIleSerIleSerIleSerIle 60
DB 1841 ACCCTGACATGTTGT 1900
QY 61 TYrTYrLeuSerSerProGluApSerProIleCyAlaTYrThLysGluPheTrpAlaTr 80
DB 1901 TACAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1960
QY 81 IleLeuApSerAlaGlyProleuProAspTYrTrpProGluIlePheSerPheGlu 100
DB 1961 ATGATGAGGACGACCAATGTCCTCCAGACCTAATGTCCTCCAGATCTTCCATTCAC 2020
QY 101 SerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLysProHisAspLeuGluProGly 120
DB 2021 ACCCGTGCAGAGTGCAGCTCTACGCGATGATCTTACAGCCACACCCCTGCACTGG 2080
QY 121 LysLeuTYrProThrValLeuPheIleTYrGlyGlyProGlu 134
DB 2081 AGGAAGACACCCACGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2140
QY 134
DB 2141 TCTTTAAGGATCAATATCTGCGGCTAATATACATGCGATCTTGGGCTATGCTGTG 2200
QY 134
DB 2201 GTGGATGATGATGTCGGGCTCTGTACAGCGGCGCTGTGACAGGAGGCGCTGAA 2260
QY 135
DB 2261 AATCAATGCGGACAGTGTGAGTGTGAGACCGAGTGTGAGGCTTGTGAGTGTGAG 2320
QY 152 ArgTYrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTYrGlyGlyTYr 171
DB 2321 AAGTATGCTTATGATCTTGTGAGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2380
QY 172 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValaIleAlaGlyAla 191

Db 2381 CTCCTACGATGGGCTCATCCAGCAAGCTTCAGGTACCATTCGGGGCGCT 2440
QY 192 ProvalThleutrpilPhePheTyAspThrGlyTyThrGluArgTyMetGlyHisPro 211
Db 2441 CCTGTACTGTGTGGATGGCTTATGACACAGGATACAGGAAGATGATGTCCTCC 2500
QY 212 AspGlnAsnGluGlnGlyTyTyTyLeuGlySerValAlaMetGlnAlaGluTySphPro 231
Db 2501 GAAATATACACAGCAAGGCTATGAGGCGGGCTGTGAGCCCTGATGTGAGAAAGCTGCC 2560
QY 232 SerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnValHisSphAla 251
Db 2561 AATGACCTTACCGCTGCTTATCTCCAGGCTTCCTGAGACGAAACGTTCTTC 2620
QY 252 HistHSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAspLeuGlnIle 271
Db 2621 CACACAAATTTCTGTGGTGTCCAGCTGATCCGAGCAGGAAGCATACAGCTTCGATC 2680
QY 272 TyTProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyTyGluLeuHis 291
Db 2681 TACCCAAAGAGAGATACATCCTGCTGCCGAGATCCGAGAGCATTAAGAGTGACG 2740
QY 292 LeuLeuHisTyTyLeuGlnGluAsnLeu 300
Db 2741 CTGCTGCACTTCTTGCAAGAACCTG 2767

RESULT 7 789 bp mRNA linear EST 07-MAY-2001
Bg709118 602675382F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4797668 5',
LOCUS mRNA sequence.
DEFINITION Bg709118.1 GI:13987136
ACCESSION Bg709118
VERSION Bg709118.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 789)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@dcf-remail.nih.gov
Tissue Procurement: Miklos Palokovics, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM10684 row: d column: 21
High quality sequence stop: 784.
Location/Qualifiers
1..789
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4797668"
/issue_type="hypothalamus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_96"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gcgcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTT-3',
size-selected for average insert size 2.3 kb and
normalized to Rot 5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

FEATURES
Source
BASE COUNT 207 a 163 c 187 g 232 t

ORIGIN

Alignment Scores:
Pred. No.: 4, 9e-108 Length: 789
Score: 1052.50 Matches: 208
Percent Similarity: 80.004 Conservative: 0
Best Local Similarity: 80.004 Mismatches: 2
Query Match: 62.654 Indels: 51
DB: 10 Gaps: 1

US-10-070-464-3 (1-310) x Bg709118 (1-789)

QY 77 PheTPrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyThrProGluIle 96
Db 10 TTTTGGGCGCACCATTTTGGATTGACAGGCTCCCTTCGACTATATCTCCAGAAATT 69
QY 97 PheSerPheGluSerThrThrGlyPheThrLeuTyGlyMetLeuTyTySphProHisAsp 116
Db 70 TTTCTTTTGAAGATCTACTGATTTACATTTGATAGGAGTGTCTTACAGAGCTCATGAT 129
QY 117 LeuGlnProGlyLysLysTyTyProThrValLeuPheIleTyTyGlyTyProGln----- 134
Db 130 CTACAGCTCGAAGAAATATCTTACTGATGCTGTTCATATATGTGTCTCTAGGTCAG 189
QY 134 ----- 134
Db 190 TTGGTAAATATCGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTAGCCTCTCTA 249
QY 134 ----- 134
Db 250 GGTATATGCTGTGTAGTATAGCAACAGGGATCTGTCCAGAGGCTTAAATTGAA 309
QY 136 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGln 147
Db 310 GCGGCTTAAATATATAAATGAGTCAATAGAAATTGACATCAGATCGAAGAGCTCCA 369
QY 148 TyTLeuAlaSerArgTyTyAspPheIleAspLeuAspArgValGlyIleHisGlyTTPSer 167
Db 370 TATCTAGCTTTCGATATATATTCATTTGATGATGCTGTGGGATCCAGGCTGTCTC 429
QY 168 TyTGLyGlyTyTyLeuSerIleuMetAlaLeuMetGlnAspIlePheArgValAla 187
Db 430 TATGAGAGATACCTCTCCCTGATGATGATATGACAGAGGTCAGATATCTTCAGGGTTGCT 489
QY 188 IleAlaGlyAlaProValThrLeuTrpIlePheTyAspThrGlyTyThrGluArgTyTyr 207
Db 490 ATTGCTGGGGCCCCCACTCACTCTGTGATCTTATGATACAGATACAGGAACGTTAT 549
QY 208 MetGlyHisProAspGlnAsnGluGlnGlyTyTyTyLeuGlySerValAlaMetGlnAla 227
Db 550 ATGGGTACCCCTGACCTGAAATGAAACAGGCTATTACTTAGATCTGTGGCATTGCAAGCA 609
QY 228 GluTySphProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsn 247
Db 610 GAAAGTTTCCCTCTTACCAATTCGTTTACTCTCTTACATGCTTTCCTGATGAGAAAT 669
QY 248 ValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTy 267
Db 670 GTCCATTTTGCACATACAGATATATTACTGAG- TTTTATGAGGGGCTGGAAGAGCATAT 728
QY 267 rAspLeuGlnIleTyTyProGlnGluArgHisSerIleArgValProGluSerGlyGlu 286
Db 729 TGATTTACAGATCTATCTCAGAGAGACACAGCATTAAGAGTCCCTGAATCGGAGAA 786

RESULT 8 957 bp mRNA linear EST 15-JUL-2002
Bg675006 AGENCOURT 9303592 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6274622
LOCUS 5', mRNA sequence.
DEFINITION Bg675006
ACCESSION Bg675006
VERSION Bg675006.1 GI:21785840
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 957)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished
	Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .957

BASE COUNT	257 a	206 c	228 g	266 t	-
ORIGIN					

Alignment Scores:		
Pred. No.:	4,81e-106	957
Score:	1036.00	Matches: 200
Percent Similarity:	76.81%	Conservative: 2
Best Local Similarity:	76.05%	Mismatches: 5
Query Match:	61.67%	Indels: 52
DB:	13	Gaps: 1
DS-10-070-464-3 (1-310)	x BQ675006 (1-957)	.

(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 280 a 227 c 242 g 292 t

ALIGNMENT SCORES:

Alignment Scores:	Length:	Matches:	Mismatches:	Indels:	Gaps:
Pred. No.:	2,376-104	1041	216	6	4
Score:	1021.50	Conservative:	6		
Percent Similarity:	73.03%	Mismatches:	16		
Best Local Similarity:	71.05%	Indels:	66		
Query Match:	60.80%	Gaps:	4		

US-10-070-464-3 (1-310) x B1084090 (1-1041)

QY 1 PHEGLUGLThryAspSerProLeuGluHisLeuTyrrValValSerTyrrValAsn 20
 DB 153 TTTGAAGGCAACCAAACTCCCTTTAGAGCATCAGTCTAGTTCAGTTACGTTAAAT 212
 QY 21 PROGLGLUValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 40
 DB 213 CCGAGAGAGTGACAAAGCGTGAACCGGCTACTACATTCCTGTCGACATCAG 272
 QY 41 HisCysAspPhePheIleSerIlyrSerAsnGlnLysAsnProHisCysValSerLeu 60
 DB 273 CACTGAGCATCTCTTTAAGTAGTAGTAGTACACAGACATCACACTGTGTCTCTT 332
 QY 61 TyrrLeuSerSerProGluAspAspProThrCysLeuThrIlyGluPheThrPalThr 80
 DB 333 TACAACTTACAACTCTGAGAGTGAACCCAACTGCAAAACAAAGAAATTTTGCGCCACC 392
 QY 81 IleLeuAspSerIleGlyProLeuProAspTyrrThrProGluIlePheSerPheGlu 100
 DB 393 ATTTGAGATTCAGCAGATCCTCTCTGACTATACCTCCCAAAATTTTCTTTTGA 452
 QY 101 SerThrThrGlyPheThrLeuTyrrGlyMetLeuTyrrLysProHisAspLeuGlnProGly 120
 DB 453 AGTACTACTGAGATTTCATTTGATTTGGAGTGCCTTACAGAGCTCATGATCTACAGCCTGGA 512
 QY 121 LysIlyrTyrrProThrAlaLeuPheIleTyrrGlyIlyProGln 134
 DB 513 AAGAAATATCTACTGTGCTGCTCATATATGAGTGCTCCTCAGGTGAGTGAATTAAT 572
 QY 134 573 CGGTTTACAGAGAGTCAAGTATTTCGGCTGAATACCTAGCCTCTAGATTATGTGG 632
 DB 134 134
 QY 134 TTTAGTATGATAGCAACAGGGGATCTGTCCAGGAGGCTTAATTGAAGGCGCTTTA 692
 DB 633 TTTAGTATGATAGCAACAGGGGATCTGTCCAGGAGGCTTAATTGAAGGCGCTTTA 692
 QY 135 693 ATATACAAATGGGTCATATGAATATGACGATCAGTGGAGAGACTCCAAATTTATTTACCTTC 752
 DB 693 ATATACAAATGGGTCATATGAATATGACGATCAGTGGAGAGACTCCAAATTTATTTACCTTC 752
 QY 151 rArgTyrrAspPheIle-AspLeuAspArgValGlyIleHisGlyTyrPseSerTyrrGly-Gly 170
 DB 753 TCGATATGATTTCAATTGACTTATGATTCGCTGGGCAATCCAGGGTGTCTTATGAGAGGA 812
 QY 171 TyrrLeuSerLeu-MetAlaLeuMetGlnArgSerAsp-IlePheArgValAlaIleAla- 189
 DB 813 TACCTCTCTCCATATGAGCATTAATGAGAGGTAGATGAGTTCAGGGGTCATATGAGT 872
 QY 190 GlyAlaProValThrLeuTrpIlePhe-TyrrAspThrGlyTyrrThrGluArgTyrrMetGln 209
 DB 873 GGGGGCCCACTGATCTGTGATCTTCTATATGATACAGATACAGGAAGCTTATATGGG 932
 QY 209 yHisProAspGlnAsnGlnGlnGlyTyrrTyrrLeuGlySerValAlaMetGlnAlaGly 229
 DB 933 TCACCTGAGACAGA---TGAACGGGTCAATCTTCAGGATCTGTG---CCTGACAGCAA 983
 QY 229 sPhePro-----SerGluProAsnArgLeuLeuLeuLeuHis 241

DB 984 ATTCCTCTGAGACATGCTTATATGCTTCATATGTTCCGGGCAATTCCTTCGACATACAC 1041

RESULT 10
 LOCUS
 DEFINITION

BY751026 746 bp mRNA linear EST 17-DEC-2002
 BY751026 RIKEN full-length enriched, activated spleen Mus musculus
 cDNA clone F830018D07 5', mRNA sequence.

ACCESSION
 BY751026
 BY751026.1 GI:27181193

KEYWORDS
 EST.
 Mus musculus (house mouse)

SOURCE
 ORGANISM

REFERENCE
 AUTHORS

1 (bases 1 to 746)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Oshino, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,
 Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,
 Geisel, K. M., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,
 L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,
 A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,
 Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,
 King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pettes, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,
 Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J. A., Ring,
 B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou,
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,
 R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L. G., Wynchaw-Boris, A., Yanagisawa,
 M., Yang, T., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Konno, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
 K., Shinagawa, A., Yasunishi, A., Yoshino, N., Waterston, R., Lander,
 E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

JOURNAL
 MEDLINE
 PUBMED

Nature 420, 563-573 (2002)
 2254683
 12466851

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-researc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,
 S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
 Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,
 H., Koyu, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
 Numazaki, R., Ohno, M., Oshino, N., Saito, R., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,
 Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct
 Submision

Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1571-1571 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/RRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source 1..746
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="taxon:10090"
 /clone="F83001BD07"
 /tissue_type="activated spleen"
 /clone_lib="RIKEN full-length enriched, activated spleen"
 BASE COUNT 194 a 178 c 176 g 197 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 9,386-102 Length: 746
 Score: 996.50 Matches: 189
 Percent Similarity: 77.73% Conservative: 3
 Best Local Similarity: 76.52% Mismatches: 5
 Query Match: 59.32% Indels: 50
 DB: 14 Gaps: 1

US-10-070-464-3 (1-310) x BY751026 (1-746)

Qy 17 SerTyrValAsnProGlyGlyValThrArgLeuThrAspArgGlyTyrSerHisSerCys 36
 Db 1 AGTTATGCAAAACCTCGAGAGAGTGGAGAGCTGAGACCGTGGCTACTACACACTCTGC 60
 Qy 37 CysHisSerGlnHisCysAspPhePheHisSerIleTyrSerAsnGlnValAsnProHis 56
 Db 61 TGCCACACCGGCAATGGATCTTCTTCATAGTAAAGTACAGCAACAGAGAAATCCACAC 120
 Qy 57 CysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysIleThrIleGlu 76
 Db 121 TGTGTCTCTCTTCAAACTCTCAAGTCTCTGAGATGACCCGTTCAATAAACAAGGAA 180
 Qy 77 PheTPrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProGluIle 96
 Db 181 TTTTGGGCACCATTTTGGATTACAGAGTCTCTTCTGACTACACCCCTCCAGAAATT 240
 Qy 97 PheSerPheGluSerThrThrGlyPheThrIleTyrGlyMetLeuTyrIleProHisAsp 116
 Db 241 TTTTCTTTTGAAGACTACTGATTTTACACTGATGAGATGTTGATTAAGCCTCATGAC 300
 Qy 117 LeuGlnProGlyIleValTyrProThrValLeuPheIleTyrGlyGlyProGln----- 134
 Db 301 CTACAACTCGAAAGAAATACCCCACTGTGTATTCATATATGATGATGCCAGGTGCAG 360
 Qy 134 ----- 134
 Db 361 CTGGTGAACAATCGGTTTAAAGAGTCAAGTATTCGGCTGGAACACCTGCGCTCCCTG 420
 Qy 134 ----- 134
 Db 421 GGTATATGTTGGTGTGATAGACAAGGGATCTGTCAACGAGACTTAATTTGAA 480
 Qy 135 -----GlyGlnIleGluIleAspAspGlnValGluGlyLeuGln 147
 Db 481 GGCCTCTTAATATATTAATGGTCAATAGAAATCGATGATCAAGTGAAGAGACTCCAG 540
 Qy 148 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTyrPhe 167
 |||||||.....

Db 541 TACCTAGACATCTCAGTATGACTTTCATTGATCGATCGAGGATCCACGCTGCTCC 600
 Qy 168 TyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAspSerAspIle-PheArgValAl 187
 Db 601 TATGTGTGTACTCTCTCTCTGATGACATTAAAGCAGAGGTGCAATATCTTCCGGGTGC 660
 Qy 187 allelaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGluArgTyr 207
 Db 661 TATTCTGGGGGCCAGTACACCTCTGTGATCTTCTATATATACAGATACAGAGGCGCTA 720
 Qy 207 rNeGlyHisProAspGln 213
 Db 721 TATGGTCAACCTGACACAG 739

RESULT 11
 B0675260 910 bp mRNA linear EST 15-JUL-2002
 LOCUS AGENCOURT 8354972 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275478
 DEFINITION 5', mRNA sequence.
 ACCESSION B0675260
 VERSION B0675260.1 GI:21786094
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 910)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strassberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
<http://image.llnl.gov>
 Plate: L1CM2458 row: 1 column: 07
 High quality sequence stop: 618.

FEATURES

source 1..910
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6275478"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="RDH10B (phage-resistant)"
 /clone_lib="NIH_MGC_102"
 /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAGG). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."
 BASE COUNT 248 a 186 c 212 g 262 t 2 others
 ORIGIN

Alignment Scores:

Pred. No.: 5,33e-101 Length: 910
 Score: 991.00 Matches: 195
 Percent Similarity: 77.25% Conservative: 2
 Best Local Similarity: 76.47% Mismatches: 53
 Query Match: 58.99% Indels: 5
 DB: 13 Gaps: 2

US-10-070-464-3 (1-310) x B0675260 (1-910)

Qy 1 PheGluGlyThrIleAspSerProLeuGluHisIleLeuTyrValIleSerTyrValAsn 20
 |||||||.....

Db 605 CATGGGTTCCGTGATGAGATGTCATTTTGCACATACCATATATTACGAGTTT 664
 Qy 260 euVal-ArgAlaGly-LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSer11 279
 Db 665 TAGTGGAGGGCTGGAGAAACCATATATATTACGATCTATCCCTCAGAGACACAGCAT 724
 Qy 279 eArgValProGluSerGlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAs 299
 Db 725 AAGAGC-CCTGAATCGGAGAACTTATGACATGCTCTTTGGCC--TACCTTCAGAAAA 780
 Qy 299 nleu 300
 Db 781 CCTG 784
 RESULT 13
 BU447277 735 bp mRNA linear EST 29-NOV-2002
 LOCUS 60376685F1 CSEOREN13 Gallus gallus CDNA Chest693j22 5', mRNA
 DEFINITION BU447277
 ACCESION BU447277
 VERSION BU447277.1 GI:25936588
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 735)
 Boaden, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22355534
 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 FEATURES
 source 1..735
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Layer"
 /db_xref="taxon:9031"
 /clone="CHEST693j22"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEOREN13"
 /note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. CDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was bluntended, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

BASE COUNT 183 a 170 c 186 g 196 t
 ORIGIN
 Alignment Scores: 8.17e-98 Length: 735
 Pred. No.:

Score: 961.50 Matches: 186
 Percent Similarity: 78.694 Conservative: 6
 Best Local Similarity: 76.234 Mismatches: 2
 Query Match: 57.234 Indels: 50
 DB: 13 Gaps: 1
 US-10-070-464-3 (1-310) x BU447277 (1-735)
 Qy 110 MetLeuTyrLysProHisAspLeuGlnProGlyLysValTyrProThrValLeuPhe11e 129
 Db 6 ATGATGTACAAACCTTCACATCTGCACAACTGGAAAAAGATACCTTACGTGATCTTCATC 65
 Qy 130 TyrGlyGlyProGln----- 134
 Db 66 TATGAGGCCCTCAGGTGACGCTAGTGAACAATTCGATTTAAAGACTCAAAATTTCCGA 125
 Qy 134 ----- 134
 Db 126 TTGAACACCTTGACCTTTAGGCTATGTTGTTGTTATTGCAACACGGGGCTCTGC 185
 Qy 135 -----GlyGlnIleGlu1leap 140
 Db 186 CACCGAGGCTGAAGTTTGAGAGAGCCTTAAATACAAATAGGACAAATAGAAATTGAT 245
 Qy 161 AepGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPhe1leAspLeuAspArg 160
 Db 246 GACCAAGTGAAAGGCTGCACTACTTAGGCTCCGCTGATGACTTCTTATTTGGATCGT 305
 Qy 161 ValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180
 Db 306 GTTGGCATTCAGGGCTGGTCCATAGGAGCTACCTCTCTTATGCTTTAATGACAGAG 365
 Qy 181 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 200
 Db 366 TCAGATATCTTCAGGGTTGCGATCGCGGAGACCCGTCACCTGGAGATTTTCTATGAC 425
 Qy 201 ThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAspGlnGlnIleTyrTyrLeu 220
 Db 426 ACGGGTTACACGAGAGCGCTTACATGAGCCACCGGACACAGACAGAGGACTTATTAACCTG 485
 Qy 221 GlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArg-LeuLeuLeuLe 240
 Db 486 GGTTCAGTGGCCATCAGCTAGAGATTCTCTTGCAGAACCAACCGTTTGTCTGCTGCT 545
 Qy 240 uHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSer1leLeuLeuSerPheLe 260
 Db 546 ACATGGGTTCTTGGATGAGAACGTTCACTTGCACACACTAGTATTTTCTCAGCTTTT 605
 Qy 260 uValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgHisSer1leAr 280
 Db 606 AGTGAAGAGCTGGAGAACCTACAGACTTGCAGATCTACCTCAGAGAGGACACAGTATAG 665
 Qy 280 gValProGluSerGlyGlyHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLe 300
 Db 666 GGTGCTGAGTGGGAGAGCACTATGAATCCTCATCTAGTATTTACCTGCAAGAGATCT 725
 Qy 300 uGlySerArg 303
 Db 726 GGGCTCTCGC 735
 RESULT 14
 BI084885/c 855 bp mRNA linear EST 20-JUN-2001
 LOCUS 602869453T1 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:5013996 3',
 DEFINITION mRNA sequence.
 ACCESION BI084885
 VERSION BI084885.1 GI:14503215
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 855)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONETECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LNCM1820 row: j column: 13
 High quality sequence start: 3
 High quality sequence stop: 844.
 Location/Qualifiers

FEATURES

source

1..855
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5013996"
 /issue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH MGC 102"
 /note="Torgan: salivary gland; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-CDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies). Note: this is a NIH-MGC Library."
 BASE COUNT 244 a 200 c 172 g 239 t
 ORIGIN

Alignment Scores:

Pred. No.: 3,14e-95 Length: 855
 Score: 939.50 Matches: 196
 Percent Similarity: 77.738 Conservative: 3
 Best Local Similarity: 76.568 Mismatches: 3
 Query Match: 55.924 Indels: 55
 DB: 12 Gaps: 1

US-10-070-464-3 (1-310) x BI084885 (1-855)

Oy 109 GlyMetLeuTyrLysProHisAsp-LeuGlnPro-GlyLysLysTyrProThrValLeuP 128
 Db 854 GGGATGCTCTACAGAGCTCAGAGCTCTACAGCTGGGAAAGATATCTTACTGTCTGT 795
 Oy 128 heileTyrGlyLysProGln----- 134
 Db 794 TCATAAATGTGTGCTCTCA-GGTGCAATTGTGTAATATCGGTTTAAAGAGTCAATAT 736
 Oy 134 ----- 134
 Db 735 TTCGGTTGAATACCTAGCCTCTAGGTTATGCGTTGTAGTATAGACAACAGGGA 676
 Oy 135 -----GlyGlnIleG 138
 Db 675 TCCTGTCACGAGGGCTTAATATTGAAGGCGCTTAATATTAATAGGTCATAATAG 616
 Oy 138 LuLeAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAsp-PheIleAsp 157
 Db 615 AAATTGAGATCAGTGGAGAGCTCCATATCTAGCTTCTGATATGAGTTTCATGAC 556
 Oy 158 LeuAspArgValGlyIleHisGlyTyrSerTyrGlyGlyTyrLeuSerLeuMetAlaLeu 177
 Db 555 TTAGATCGTGTGGGATCCACGGCTGTGCTCTATGAGAGATACCTCCCTGATGGCATTA 496
 Oy 178 -MetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTyrI 197
 Db 495 ACTGAGAGGTCAGATATCTTCAAGGTTGCTATGCTGGGCGCCAGTCACTTGTGAT 436

Oy 190 ePheTyrAspThrArgIleTyrThrGlnArgTyrMetGlyHisProAspGlnAsnGlnGln 217
 Db 435 CTTTATGATATACAGATACACGGAACGTTATGCGTCACTGCACCAATGAACAGCG 376
 Oy 217 YTTTyrLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAspArgLe 237
 Db 376 CTATTACTTAAAGATCTGTGGCCATGCAAGCAAAAAAGTTCCCTCGAACCAAAATGTT 316
 Oy 237 uLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 257
 Db 315 ACTGCTCTTACATGATGTTCTCTGATGAGATATCCATTTTGACATACAGATATTA 256
 Oy 257 uSerPheLeuValAlaGlyAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArgH 277
 Db 255 GAGTTTATAGAGAGGCTGGAAGCAACCATATGATTAAGATCTATCTCAGAGAGACA 196
 Oy 277 sSerIleArgValAlaProGluSerGlyGlnHisTyrGluLeuHisLeuHisIleLeu 297
 Db 195 CAGCATTAAGAGTTCCTGAATCGGAGAACATTATGAACCTGATCTTTGCACTACCTTCA 136
 Oy 297 nGluAsnLeuGlySerArgIleAlaAlaLeuIleValIle 310
 Db 135 AGAAAACTTGATCACGATATGCTGCTCTAAAGTGA 96

RESULT 15
 AL582206 1085 bp mRNA linear EST 01-JUN-2003
 LOCUS AL582206 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
 DEFINITION Homo sapiens CDNA clone CS0DL005YD02 3-PRIME, mRNA sequence.
 ACCESSION AL582206
 VERSION AL582206.2 GI:31320424
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1085)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length CDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT On Feb 16, 2001 this sequence version replaced gi:12949961.

Contract: Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 7542.r for
 more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DL005DB01NP1&cluster=7542.r>. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>
 Faraday Avenue Genoscope sequence ID : CS0DL005DB01NP1.

FEATURES

source

1..1085
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DL005YD02"
 /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
 /cell_line="RAMOS CELL LINE"
 /clone_id="RAMOS CELL LINE"
 /note="1st strand CDNA was primed with a NotI-oligo (dT)
 primer. Five prime end enriched, double-strand CDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 BASE COUNT 297 a 221 c 220 g 315 t 32 others
 ORIGIN

Alignment Scores:

Pred. No.: 6.72e-95 Length: 1085
 Score: 938.00 Matches: 221

Percent Similarity: 70.40%
Best Local Similarity: 68.85%
Query Match: 55.83%
DB: 9
Conservative: 5
Matches: 25
Indels: 71
Gaps: 5

US-10-070-464-3 (1-310) x AL582206 (1-1085)

```
OY 2 GIUGLYThrLysAspSerProLeuGluHisIleuTyrValIleSerTyrValIleAspPro 21
DB 1021 GAAGGACACCAAGATTCCTCCCTTACGAGMATCCTCTACGTCGAGMATTCCTTACCT 962
OY 22 GIGLYValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGlnHis 41
DB 961 GGAGAGGTGAMAGGCTGACTGACCGTGTACTCATCTTCGTCGATTCAGTCAGCAC 902
OY 42 CysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyr 61
DB 901 TGTGACTTCTTATAAKTAAGTATAGTAAACGAAAGATCCACACGTGTGCTCCCTTAC 842
OY 62 LysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIle 81
DB 841 AAGCTATCAAGTCCGAAATGACCCAACTTGCAAAACAAAGAAATTTGGGCTACCAT 782
OY 82 LeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSer 101
DB 781 TTGGATTTCAGCAGTCTCTCTCTGACTACTCTCCAGAAATTTCTCTTTGAAAGT 722
OY 102 ThrThrGlyPheThrIleuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLys 121
DB 721 ACTACGAGATTACATGATGAGATGCTCTACAAAGCTCATGATCTACAGCTCGAAAG 662
OY 122 LysTyr-ProThrValLeuPheIle-TyrGlyGly-ProGlnGlyIleGluIleAsp 140
DB 661 ATATATCKTACTGTGCTCTCTCATATATGAGTGCTGCGAGGTAGTA----- 611
OY 141 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 160
DB 611 ----- 611
OY 161 ValGlyIleHisGlyTyrPheTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 180
DB 611 ----- 611
OY 181 SerAspIlePheArgValAlaIleAlaGlyAla-ProValThr-LeuTrpIlePheTyr- 199
DB 610 -----TTTGCTGGGGCCCCAGTCACATCTGTGATCATTTACTG 572
OY 200 --AspThrGlyTyrThrGluArg-TyrMetGly-HisProAspGlnAsn-GluGlnGly- 217
DB 571 ATGATACAGGATACACGGAACGATTATATGCGCTCACCTGACCAAGATAGAACAGGGAC 512
OY 218 TyrTyrLeuGly-SerValAlaMetGlnAlaGluLys-PheProSerGluProAsnArgL 237
DB 511 TATTACTTAGGATGCTGTGGCCATGCAACAGAAATAGTACCCCTCTGAAACCAATCGTT 452
OY 237 eulLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 257
DB 451 TACTGCTCTTACATGATGTTCCGATGAGATGTCATTTGTCACATACCATATATTAC 392
OY 257 eulSerPheLeuValArg--AlaGlyLysProTyr--AspLeuGlnIleTyrProGlnG 275
DB 391 TGAGTTTTTTTAGTCGAGTCCGCTGGAAGCG-CATATGCAATTTA-----CAG 345
OY 275 IuArgHisSerIleArgValProGluSerGlyLysHisTyrGluLeuHisIleLeuHis- 294
DB 344 AAGACACAGCATTAAGATTCCTGAAATCGGAAACATTATGAACCTGCACTTTTGCACT 285
OY 295 -TyrLeuGlnGluAsnLeuGlySerArgIleAlaAla 306
DB 284 GACTCTTCAAGAAACCTATGATGACGATATTAGCT 248
```

Job time : 1876.46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 15:44:49 : Search time 28.3565 Seconds
(without alignments)
2602.858 Million cell updates/sec

Title: US-10-070-464-5

Perfect score: 2482
Sequence: 1 TGTANPKVTFKMSEIMIDAE.....HLHLVQLQENLGRIRALKVI 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	465	22	Human DPP8 318Thr-
2	2482	100.0	724	23	Novel human protei
3	2482	100.0	782	23	Novel human protei
4	2441.5	98.4	632	22	Human protein sequ
5	2442	97.6	882	22	Human DPP8. Homo
6	2442	97.6	882	23	Human dipeptidyl p
7	2442	97.6	882	23	Human DPP4 relate
8	2442	97.6	882	23	Human protease PRT
9	2442	97.6	882	23	Amino acid sequenc

10	2422	97.6	882	24	ABU07720	Human serine prote
11	2244.5	90.4	580	23	AAE14337	Human protease PRT
12	1809.5	72.9	690	23	ABG61594	Human DPP-1 splic
13	1808	72.8	658	23	ABG61600	Human DPP-1 splic
14	1808	72.8	661	23	ABG61596	Human DPP-1 splic
15	1540	62.0	613	23	ABG61601	Human DPP-1 splic
16	1494.5	60.2	830	23	AAE24171	Human dipeptidyl p
17	1494.5	60.2	863	23	ABG61592	Human DPP4 relate
18	1494.5	60.2	882	23	ABG61593	Human DPP4 relate
19	1494.5	60.2	882	23	ABG61602	Human DPP4 relate
20	1494.5	60.2	882	23	ABG61604	Human DPP4 relate
21	1494.5	60.2	882	23	AAE24168	Human dipeptidyl p
22	1488.5	60.0	869	23	AAE24169	Human dipeptidyl p
23	1445	58.2	829	23	ABG61607	Human DPP-2 splic
24	1445	58.2	879	23	ABG61608	Human DPP-2 splic
25	1377.5	55.5	310	22	ABG61718	Human DPP8 524phe-
26	1377.5	55.5	310	23	ABG61718	Human dipeptidyl p
27	1369	55.2	847	23	AAE23875	Murine dipeptidyl
28	1298.5	52.3	497	23	ABG64845	Human albumin fusi
29	1298.5	52.3	497	23	AAU96169	Human secreted pro
30	1296	52.2	663	22	AAH40510	Human polypeptide
31	1296	52.2	737	22	AAH38724	Human polypeptide
32	1294.5	52.2	497	23	ABG64844	Human albumin fusi
33	1294.5	52.2	497	23	AAU96192	Human secreted pro
34	1187.5	47.8	508	21	AAH42928	Human ORFX ORF2692
35	1094.5	44.1	832	23	ABG61605	Human DPP-2 splic
36	1094.5	44.1	832	23	ABG61606	Human DPP-2 splic
37	1045	42.1	819	23	ABG61609	Human DPP-2 splic
38	1045	42.1	819	23	ABG61610	Human DPP-2 splic
39	1008.5	40.6	720	21	AAH41626	Human ORFX ORF1390
40	860.5	34.7	360	22	AAH47190	Human DPP8 244Glu-
41	845	34.0	1042	22	ABG60137	Drosophila melanog
42	845	34.0	1102	22	ABG62029	Drosophila melanog
43	744	30.0	250	23	ABG99949	Dipeptidyl peptida
44	663	26.7	136	22	AAH74673	Human protease and
45	663	26.7	193	23	ABH89739	Human polypeptide

ALIGNMENTS

RESULT 1	AAH47189	standard; Protein; 465 AA.
ID	AAH47189;	
AC	AAH47189;	
XX		
DT	29-JUN-2001 (first entry)	
XX		
DE	Human DPP8 318Thr-658Val-759Ala-882Ile.	
XX		
KW	Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;	
KW	dipeptidyl peptidase; DPP4; T cell; cleavage; diarrhoea;	
KW	growth hormone deficiency; glucose level; mucosal regeneration;	
KW	non-insulin dependent diabetes mellitus; glucose intolerance;	
KW	immunosuppression.	
OS	Homo sapiens.	
XX		
PN	WO200119866-A1.	
XX		
PD	22-MAR-2001.	
XX		
PF	11-SEP-2000; 2000MO-AU01085.	
XX		
PR	10-SEP-1999; 99AU-0002762.	
XX		
PR	18-FEB-2000; 2000AU-0005709.	
XX		
PA	(UNSW) UNIV SYDNEY.	
XX		
PI	Abbot CA, Gorell MD;	
XX		
DR	WPI; 2001-281520/29.	

DR N-PSDB; AAC85696.
XX New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells -
XX
XX Claim 5; Page 71-72; 78pp; English.
XX
XX The sequences given in AAB47188-90 represent fragments of human
CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for
CC H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a
CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable
CC of hydrolysing the peptide bond C-terminal to Pro in each of these
CC compounds. DPP8 is homologous with human DPP4V. DPP8 is useful for
CC cleaving a substrate, and for detecting an activated T cell which
CC involves measuring the level of DPP8 gene expression in a T cell. The
CC level of DPP8 expression is detected by detecting the amount of DPP8
CC RNA in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
XX

SQ Sequence 465 AA;

Query Match 100.0%; Score 2482; DB 22; Length 465;
Best Local Similarity 100.0%; Pred. No. 6.4e-237;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPELIFGVEYIARAGWTPGKYAMSI 60
DB 1 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPELIFGVEYIARAGWTPGKYAMSI 60
QY 61 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 120
DB 61 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 120
QY 121 HVPFQSHHEEIEFIFASECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKCPIKEIA 180
DB 121 HVPFQSHHEEIEFIFASECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKCPIKEIA 180
QY 121 HVPFQSHHEEIEFIFASECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKCPIKEIA 180
DB 121 HVPFQSHHEEIEFIFASECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKCPIKEIA 180
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVYVNGEYTRLTDRCYSH 240
DB 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVYVNGEYTRLTDRCYSH 240
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVYVNGEYTRLTDRCYSH 240
DB 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVYVNGEYTRLTDRCYSH 240
QY 241 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPTCKTKEFWATILLDSAGPLPDVTPP 300
DB 241 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPTCKTKEFWATILLDSAGPLPDVTPP 300
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVAIAGAPVTLMIYDTGYTE 360
DB 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVAIAGAPVTLMIYDTGYTE 360
QY 361 RYMGHPDQNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVFAHTSILLSLVLRAGK 420
DB 361 RYMGHPDQNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVFAHTSILLSLVLRAGK 420
QY 421 PYLQLOIYPOBRHSIRVPSGHEHYELHLHYLOENLGSRIAAKVI 465
DB 421 PYLQLOIYPOBRHSIRVPSGHEHYELHLHYLOENLGSRIAAKVI 465

RESULT 2
ID ABB97362 standard; Protein: 724 AA.
XX ABB97362;
XX AC ABB97362;
XX DT 27-JUN-2002 (first entry)
XX Novel human protein seq ID NO: 630.

XX Human, Antianemic; vulnerary; antiinflammatory; immunomodulator;
KM antineoplastic; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
KM neuroprotective; antiparkinsonian; protein therapy; EST;
KM expressed sequence tag.
XX
XX Homo sapiens.
OS
XX
XX WO200222660-A2.
PN
XX
XX 21-MAR-2002.
PD
XX
XX 10-SEP-2001; 2001WO-US26015.
PX
XX
XX 11-SEP-2000; 2000US-0659671.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Zhou P, Aundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI: 2002-292408/33.
DR
XX
XX N-PSDB; ABB97362.

An isolated polynucleotide for treating diseases associated with its
encoded polypeptide such as cancer and multiple sclerosis -
XX
XX Example 2; SEQ ID NO 630; 509pp; English.
XX

CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX
XX

Query Match 100.0%; Score 2482; DB 23; Length 724;
Best Local Similarity 100.0%; Pred. No. 1.3e-236;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPELIFGVEYIARAGWTPGKYAMSI 60
DB 260 TGTANPKVTFKMSIMIDAEGRILIDVIDKELIQPELIFGVEYIARAGWTPGKYAMSI 319
QY 61 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 120
DB 320 LDRSOTRLOIVLISPELFIPEDDVNERORLIESVDSVTPPLIYEETDININHDIF 379
QY 121 HVPFQSHHEEIEFIFASECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKCPIKEIA 180
DB 380 HVPFQSHHEEIEFIFASECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKCPIKEIA 439
QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVYVNGEYTRLTDRCYSH 240
DB 440 ITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHHLVYVYVNGEYTRLTDRCYSH 499
QY 241 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPTCKTKEFWATILLDSAGPLPDVTPP 300
DB 500 SCCISQCHDFPISKYNOKNPHCVSLYKSSPEDDPTCKTKEFWATILLDSAGPLPDVTPP 559
QY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVAIAGAPVTLMIYDTGYTE 360
DB 560 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPOVAIAGAPVTLMIYDTGYTE 619
QY 361 RYMGHPDQNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVFAHTSILLSLVLRAGK 420
DB 620 RYMGHPDQNEQGYLYGSVAMQAEKFPSEPNRLLLHGFIDENVFAHTSILLSLVLRAGK 679

QY 421 PYDLOIYPERHSIRVPESGEHEHLHLHYLOENLSRIAALVKYI 465
 DB 680 PYDLOIYPERHSIRVPESGEHEHLHLHYLOENLSRIAALVKYI 724

RESULT 3

ID ABB97361 standard; Protein; 782 AA.

AC ABB97361;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 629.

KM Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;
 KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 XX expressed sequence tag.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wenman T, Drmanac RT;

DR WPI; 2002-292408/33.

DR N-PSDB; ABN32547.

PT An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis -

PS Example 2; SEQ ID NO 629; 509pp; English.

CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to help tissue regrowth
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 782 AA;

QY Query Match 100.0%; Score 2482; DB 23; Length 782;

DB Best Local Similarity 100.0%; Pred. No. 1.5e-236;

Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKTFFKMSIEMIDAGRIIDVIDELIOPFELFEGVYIRAGTPEGKAWMSI 60
 DB 318 TGTANPKTFFKMSIEMIDAGRIIDVIDELIOPFELFEGVYIRAGTPEGKAWMSI 377
 QY 61 LIDRSQTRQIVLISPELFIPVEDDVMERQRLIESVPSVTPLIYEETDININIDIF 120
 DB 378 LIDRSQTRQIVLISPELFIPVEDDVMERQRLIESVPSVTPLIYEETDININIDIF 437
 QY 121 HVFPSHSEIEIFIPASECKTGRHLKYITSLIKESKYRSSGGLPAPDFCKPIKEEA 180
 DB 438 HVFPSHSEIEIFIPASECKTGRHLKYITSLIKESKYRSSGGLPAPDFCKPIKEEA 497

QY 181 ITSGEMVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVSIVNPGEVTRLTRDGYSH 240
 DB 498 ITSGEMVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLVYVSIVNPGEVTRLTRDGYSH 557
 QY 241 SCCISQHCDFPISKYSNQNPHCVSLYKLSPPEDPTCKTKEFWATLDSAGPLPDYTPP 300
 DB 558 SCCISQHCDFPISKYSNQNPHCVSLYKLSPPEDPTCKTKEFWATLDSAGPLPDYTPP 617
 QY 301 EIFSFESTGFTLYGMLYKPHDLQPGKYPPTYVPIYGGPQVALAGAPVTLMIFPDGYTE 360
 DB 618 EIFSFESTGFTLYGMLYKPHDLQPGKYPPTYVPIYGGPQVALAGAPVTLMIFPDGYTE 677
 QY 361 RYMGHPDQNEQGYLLASVMAQAEKFPSEPNRLLLHGFIDENYHFAHTSILSLFLVRAGK 420
 DB 678 RYMGHPDQNEQGYLLASVMAQAEKFPSEPNRLLLHGFIDENYHFAHTSILSLFLVRAGK 737
 QY 421 PYDLOIYPERHSIRVPESGEHEHLHLHYLOENLSRIAALVKYI 465
 DB 738 PYDLOIYPERHSIRVPESGEHEHLHLHYLOENLSRIAALVKYI 782

RESULT 4

ID AAB93565 standard; Protein; 632 AA.

AC AAB93565;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12964.

KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 8; SEQ ID 12964; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13632 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 632 AA;

Query Match 98.4%; Score 2441.5; DB 22; Length 632;
 Best Local Similarity 90.3%; Pred. No. 1.1e-232;
 Matches 464; Conservative 0; Mismatches 1; Indels 49; Gaps 1;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 60
 DB 119 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 178
 QY 61 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETTDIMINIHDF 120
 DB 179 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETTDIMINIHDF 238
 QY 121 HVPQSHHEEIEIFAFSECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKPIKEIA 180
 DB 239 HVPQSHHEEIEIFAFSECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKPIKEIA 298
 QY 181 ITSGEWVLCGRHSNIOQDEVRLVFEETKOSPLEHNLVYVSYNVGEVTRLTDRGYSH 240
 DB 299 ITSGEWVLCGRHSNIOQDEVRLVFEETKOSPLEHNLVYVSYNVGEVTRLTDRGYSH 358
 QY 241 SCCISQHCDFIFISKYSNQKAPHCVSILYKSSPEDDPTCKTKFMAITLDSAGLPDYTPP 300
 DB 359 SCCISQHCDFIFISKYSNQKAPHCVSILYKSSPEDDPTCKTKFMAITLDSAGLPDYTPP 418
 QY 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPQ----- 340
 DB 419 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPQVQVLRNRFKGVKRYRLNTLA 478
 QY 341 -----VAIAGAPTLMIFFDTGTYTERYMGHPONEQ 371
 DB 479 SLGIVVVVVDNRGSSCHRGKLEBGAFFKXKVAIAGAPTLMIFFDTGTYTERYMGHPONEQ 538
 QY 372 GYVLGSVAAQAEKFPSEPNRLLLHGFLDENVHFAHTSILLSFLVBAKRPYDQIYPOER 431
 DB 539 GYVLGSVAAQAEKFPSEPNRLLLHGFLDENVHFAHTSILLSFLVBAKRPYDQIYPOER 598
 QY 432 HSIRVPESGEYELHLHYLQENLGSRIALKVI 465
 DB 599 HSIRVPESGEYELHLHYLQENLGSRIALKVI 632

RESULT 5
 AAB47187 standard; Protein; 882 AA.

AC AAB47187;

DT 29-JUN-2001 (first entry)

DE Human DPP8.

XX Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;

KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;

KW growth hormone deficiency; glucose level; mucosal regeneration;

KW non-insulin dependent diabetes mellitus; glucose intolerance;

XX immunosuppression.

XX Homo sapiens.

OS

XX

Key Location/Qualifiers

FT Active-site 739
 FT /note= "Forms part of Ser-Asp-His catalytic triad"
 FT Active-site 817
 FT /note= "Forms part of Ser-Asp-His catalytic triad"
 FT Active-site 839
 FT /note= "Forms part of Ser-Asp-His catalytic triad"
 PN MO200119866-A1.
 XX
 XX 22-MAR-2001.
 XX
 XX 11-SEP-2000; 2000MO-AU01085.
 XX
 XX 10-SEP-1999; 99AU-0002762.
 PR 18-FEB-2000; 2000AU-0005709.
 XX
 XX (UNSY) UNIV SYDNEY.
 PA Abbot CA, Gorell MD;
 PI
 XX
 XX WPI; 2001-281520/29.
 DR N-PDB; AAC85694.
 XX
 PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
 PT substrates, identifying inhibitors of DPP8 catalytic activity which
 PT have therapeutic uses, and for detecting activated T cells
 XX
 PS Claim 1; Fig 2; 78pp; English.

CC This sequence represents human dipeptidyl aminopeptidase (DPP8).
 CC DPP8 has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and
 CC H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a
 CC dipeptidyl peptidase, because it is capable of hydrolysing the
 CC peptide bond C-terminal to Pro in each of these compounds. DPP8
 CC is homologous with human DPPIV. DPP8 is useful for cleaving a
 CC substrate, and for detecting an activated T cell which involves
 CC measuring the level of DPP8 gene expression in a T cell. The level
 CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
 CC in the cell. It is also useful for identifying a molecule capable
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules
 CC identified as inhibiting DPP8 catalytic activity may be useful for
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
 CC in non-insulin dependent diabetes mellitus and other disorders
 CC involving glucose intolerance, enhancing mucosal regeneration and
 CC as immunosuppressants.

SO Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 22; Length 882;
 Best Local Similarity 82.3%; Pred. No. 1.6e-230;
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 60
 DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKXAMSI 377
 QY 61 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETTDIMINIHDF 120
 DB 378 LLDRSQTRLOIVLISPELFIPVEDDVMERQRLIESVDSVTPLIYEETTDIMINIHDF 437
 QY 121 HVPQSHHEEIEIFAFSECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKPIKEIA 180
 DB 438 HVPQSHHEEIEIFAFSECKTGRHLKYKITSILKESKYKSSGGLPAPSDPKPIKEIA 497
 QY 181 ITSGEWVLCGRHSNIOQDEVRLVFEETKOSPLEHNLVYVSYNVGEVTRLTDRGYSH 240
 DB 498 ITSGEWVLCGRHSNIOQDEVRLVFEETKOSPLEHNLVYVSYNVGEVTRLTDRGYSH 557
 QY 241 SCCISQHCDFIFISKYSNQKAPHCVSILYKSSPEDDPTCKTKFMAITLDSAGLPDYTPP 300
 DB 558 SCCISQHCDFIFISKYSNQKAPHCVSILYKSSPEDDPTCKTKFMAITLDSAGLPDYTPP 617
 QY 301 EIFSFESTTGFTLYGMLYKPHDLOPGKKYPTVLFIYGGPQ----- 340

Db 618 EIFFESTTGTFLYGMLYKPHDLPQKKTPTVLFYGGPOVOLVNNRFGVKYFRLLNTLA 677
 Qy 341 ----- 340
 Db 678 SLGYVVVVINDRNGSCHRLKFEKAFYKMGQIEIDDQVEGLOYLASRYDFILDRVGIHG 737
 Qy 341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLYGSVAM 380
 Db 738 WSYGYLSLMLMQRSDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLYGSVAM 797
 Qy 381 QAEKFPSEPNRLLLHGFLENVHFAPHTSILSFVLRACKPYDLOIYQERHSIRVPESG 440
 Db 798 QAEKFPSEPNRLLLHGFLENVHFAPHTSILSFVLRACKPYDLOIYQERHSIRVPESG 857
 Qy 441 EHYELHLHYLOENLGSRIALVKVI 465
 Db 858 EHYELHLHYLOENLGSRIALVKVI 882

RESULT 6

AAE24170 ID AAE24170 standard; Protein; 882 AA.

AAE24170;

23-SEP-2002 (first entry)

Human dipeptidyl peptidase 8 (DPP8) protein.

Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
 autoimmunity; human immunodeficiency virus; HIV infection; cytostatic;
 graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
 antiviral; enzyme.

Homo sapiens.

WO200234900-A1.

02-MAY-2002.

29-OCT-2001; 2001WO-AU01388.

27-OCT-2000; 2000AU-0001076.

(UNSY) UNIV SYDNEY.

Abdott CA, Correll MD;

WPI; 2002-454646/48.

N-PSDB; AAD38956.

New dipeptidyl peptidase (DPP) peptidases, useful for screening
 inhibitors of DPP catalytic activity, which may be employed to treat
 e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 rejection and HIV infection -

Example; Fig 1; 91pp; English.

The present invention relates to dipeptidyl peptidase (DPP) proteins and
 polynucleotides encoding such proteins. The DPP peptidases are useful for
 screening inhibitors of DPP catalytic activity. The inhibitors are useful
 for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 rejection and HIV (human immunodeficiency virus) infection. The present
 sequence is human DPP8 protein.

Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 23; Length 882;

Best Local Similarity 82.3%; Pred. No. 1.6e-230;

Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

1 TGTAPKVTFKMSEIMDAEGRIIDVIDKELIQPEILFEGVEYIARAGWTFEGKYAWSI 60

Db 318 TGTAPKVTFKMSEIMDAEGRIIDVIDKELIQPEILFEGVEYIARAGWTFEGKYAWSI 377
 Qy 61 ILDRSQTRLOQIVLISPELFIPEDDVMERORLIESVPSVTBLIYEETTDIMINIDIF 120
 Db 378 ILDRSQTRLOQIVLISPELFIPEDDVMERORLIESVPSVTBLIYEETTDIMINIDIF 437
 Qy 121 HVPFQSHHEEELFIPIASECKTGFRHLKYITSILKSKYKRSSGGLPAPSDFCPIKEIA 180
 Db 438 HVPFQSHHEEELFIPIASECKTGFRHLKYITSILKSKYKRSSGGLPAPSDFCPIKEIA 497
 Qy 181 ITSGMEVLGRGNSIQVDEVRLVYFEGTKDSPLEHLYVVSYNPGCVTLTORGSYH 240
 Db 498 ITSGMEVLGRGNSIQVDEVRLVYFEGTKDSPLEHLYVVSYNPGCVTLTORGSYH 557
 Qy 241 SCCISQHCDFPISKYSNOQNPCHVSILYKLSPEDDPTCKTEFWATTILDSAGPLDPYTP 300
 Db 558 SCCISQHCDFPISKYSNOQNPCHVSILYKLSPEDDPTCKTEFWATTILDSAGPLDPYTP 617
 Qy 301 EIFFESTTGTFLYGMLYKPHDLPQKKTPTVLFYGGPO----- 340
 Db 618 EIFFESTTGTFLYGMLYKPHDLPQKKTPTVLFYGGPOVOLVNNRFGVKYFRLLNTLA 677
 Qy 341 ----- 340
 Db 678 SLGYVVVVINDRNGSCHRLKFEKAFYKMGQIEIDDQVEGLOYLASRYDFILDRVGIHG 737
 Qy 341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLYGSVAM 380
 Db 738 WSYGYLSLMLMQRSDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLYGSVAM 797
 Qy 381 QAEKFPSEPNRLLLHGFLENVHFAPHTSILSFVLRACKPYDLOIYQERHSIRVPESG 440
 Db 798 QAEKFPSEPNRLLLHGFLENVHFAPHTSILSFVLRACKPYDLOIYQERHSIRVPESG 857
 Qy 441 EHYELHLHYLOENLGSRIALVKVI 465
 Db 858 EHYELHLHYLOENLGSRIALVKVI 882

RESULT 7

ABG61591 ID ABG61591 standard; Protein; 882 AA.

ABG61591;

12-AUG-2002 (first entry)

Human DPPIV related serine protease DPP-1.

Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
 DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 heart failure; hypertension; urinary retention; osteoporosis; cancer;
 ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 dyskinesia; reproductive disorder; inflammatory disorder;
 metabolic disorder.

Homo sapiens.

WO200231134-A2.

18-APR-2002.

12-OCT-2001; 2001WO-US31874.

12-OCT-2000; 2000US-240117P.

(FERR) FERRING BV.

Qi S, Akinsanya KO, Riviere PJ, Junten J;

WPI; 2002-444176/47.

DR N-PSDB; ABK83322.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Claim 17, Fig 1, 113pp, English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, aches, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPP proteins.

XX Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 23; Length 882;
Best Local Similarity 82.3%; Pred. No. 1.6e-230;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKYTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 60
DB 318 TGTANPKYTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 377
QY 61 LLDRSQTRIQIIVISPELFIPEVDVMERORLIESVPSDSTPLIYEETTDIWINHDF 120
DB 378 LLDRSQTRIQIIVISPELFIPEVDVMERORLIESVPSDSTPLIYEETTDIWINHDF 437
QY 121 HWFPGSHHEIEIFIPASECKTGRHLKYITSILKSKYKSSGGLPAPSDPKPIKEEIA 180
DB 438 HWFPGSHHEIEIFIPASECKTGRHLKYITSILKSKYKSSGGLPAPSDPKPIKEEIA 497
QY 181 ITSGEWEVLGRHGSNIQVDEVRLLYFEGTKDPLHLLYVSVNPGVETRLTIDGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRLLYFEGTKDPLHLLYVSVNPGVETRLTIDGYSH 557
QY 241 SCCISQCHDFISKYKSNQKPNHCVSLYKLSPPEDDCTCKEPMATILDSAGLPYTPP 300
DB 558 SCCISQCHDFISKYKSNQKPNHCVSLYKLSPPEDDCTCKEPMATILDSAGLPYTPP 617
QY 301 EIFSFESTTGFTLYGMLYKPHLDQPGKKYPTVLFIYGGPQ----- 340
DB 618 EIFSFESTTGFTLYGMLYKPHLDQPGKKYPTVLFIYGGPQVOLNNRFKGVYFRNLTLA 677
QY 341 ----- 340
DB 678 SLGYVVVVINDNGSCHRGKLFBSGAFYKKGQIIFIDQVEGLQYLASRYDFIDLRVGIHG 737
QY 341 -----VALAGAPVTLMIFYDTGYTERYVGHGDQEGGYLGSVAM 380
DB 738 WSYGYVLSMALMQRSDIFRVAIAGAPVTLMIFYDTGYTERYVGHGDQEGGYLGSVAM 797
QY 381 QAEKFPSEBNRLLLHGFIDENHFAHTSLLSFLVRACKPYDLQIYPOERHSIRVPSG 440
DB 798 QAEKFPSEBNRLLLHGFIDENHFAHTSLLSFLVRACKPYDLQIYPOERHSIRVPSG 857
QY 441 EHYELHLHYLOENLGSRIALAKVI 465
DB 858 EHYELHLHYLOENLGSRIALAKVI 882

RESULT 8
AAU74749

ID AAU74749 standard; Protein; 882 AA.
XX
AC AAU74749;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 protein sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX
OS Homo sapiens.
XX
PN MO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001MO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCYTE) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM;
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hatalla A, Khan FA,
PI Wallis NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L,
PI Kallick DA;
XX
DR WPI; 2002-090437/12.
XX
DR N-PSDB; ABK12892.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 1; Page 140-142; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 882 AA;

Query Match 97.6%; Score 2422; DB 23; Length 882;
Best Local Similarity 82.3%; Pred. No. 1.6e-230;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKYTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 60
DB 318 TGTANPKYTFKMSIMIDAGRIIDVIDKELIOPFELFEGVEYIARAGTPEGKXAWSI 377
QY 61 LLDRSQTRIQIIVISPELFIPEVDVMERORLIESVPSDSTPLIYEETTDIWINHDF 120

```

Db      378 LDRSOTRLQIVLISPELFIPVEDDVMERORLIESVDSVTPLIYEETDLMINIHDF 437
Qy      121 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIMEIA 180
Db      438 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIMEIA 497
Qy      181 ITSGEWEVLGRGNSIQVDEVRRLVFEETKOSPLEHLLYVSYVNPGEVTRLTDGYSH 240
Db      498 ITSGEWEVLGRGNSIQVDEVRRLVFEETKOSPLEHLLYVSYVNPGEVTRLTDGYSH 557
Qy      241 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 300
Db      558 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 617
Qy      301 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQ----- 340
Db      618 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQVOLLVNNRFKGVKXYRLNTLA 677
Qy      341 ----- 340
Db      678 SLGYYVVVVIDNRGSGRGLKFECAFKYKMGQIETDQVEGLQYLASRYPIDLDRYING 737
Qy      341 -----VALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 380
Db      738 WSYGYLSLMLMQRSDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 797
Qy      381 QAEKFPSEBNRLLHLGFLDENVHFHTSILSLFVRACKPYDLQIYPERHSIRVPEG 440
Db      798 QAEKFPSEBNRLLHLGFLDENVHFHTSILSLFVRACKPYDLQIYPERHSIRVPEG 857
Qy      441 EHYELHLHYLOENLGSRIAALKVI 465
Db      858 EHYELHLHYLOENLGSRIAALKVI 882

RESULT 9
AAG78415
ID      AAG78415 standard; Protein; 882 AA.
XX
AC      AAG78415;
XX
DT      12-APR-2002 (first entry)
XX
DE      Amino acid sequence of 21953 human prolyl oligopeptidase.
XX
KW      21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
KW      cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW      neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW      antidiabetic; antiarthritic; antiaesthetic; antiinflammatory;
KW      diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW      Grave's disease; neuronal disorder; demyelinating disease.
XX
OS      Homo sapiens.
XX
PN      WO200179473-A2.
XX
PD      25-OCT-2001.
XX
PF      11-APR-2001; 2001WO-US40483.
XX
PR      18-APR-2000; 2000US-197508P.
XX
PA      (MILL-) MILLENNIUM PHARM INC.
XX
PI      Meyers RA, Williamson M;
XX
DR      MPI; 2002-034353/04.
XX
DR      N-PSDB; AAH99934.
XX
PT      New polypeptides 21953, member of human prolyl oligopeptidase family,
PT      useful as diagnostic targets and therapeutic agents for controlling
PT      cancer, lymphoma and leukemia
XX

```

```

PS      Claim 1; Page 102-103; 121pp; English.
XX
CC      This invention relates to an isolated 21953 human prolyl
CC      oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
CC      neuroprotective, antihypertensive, dermatological, antipsoriatic,
CC      antiepileptic, ophthalmological, antiinflammatory, nootropic,
CC      antiparasitoma, anticonvulsant, gynaecological, vasorelaxant,
CC      antitumoral, cardiant, anticholesterolic, anorectic and
CC      metabolic in its action. Uses include gene therapy, expression or
CC      activity of 21953 protein modulator, it is useful for identifying a
CC      compound which binds to it and can be used in preventing, treating
CC      or detecting a cellular proliferative or differentiative disorder.
CC      The 21953 molecules can act as novel diagnostic targets and therapeutic
CC      agents for controlling disorders associated with the aberrant activity
CC      or degradation of peptide hormones e.g., disorders associated with cell
CC      differentiation and proliferation such as cancer, immune function,
CC      reproductive, neurological and cardiovascular function. The 21953
CC      molecules are thus useful for treating and preventing cellular
CC      proliferative and differentiative disorders, haematopoietic neoplastic
CC      disorders, immune disorders such as autoimmune diseases, diabetes
CC      mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC      neuronal disorders, demyelinating diseases, vascular disorders and
CC      metabolism or pain disorders. This sequence represents the amino
CC      acid sequence of 21953 human prolyl oligopeptidase.
XX
SQ      Sequence 882 AA:

```

```

Query Match      97.6%; Score 2422; DB 23; Length 882;
Best Local Similarity 82.3%; Pred. No. 1,6e-230;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

```

```

Qy      1 TGTANPKVTFKMSIEMIDAEGRILIDYIDKELLQPFILLEGYFIRARACTPPEKAWMSI 60
Db      318 TGTANPKVTFKMSIEMIDAEGRILIDYIDKELLQPFILLEGYFIRARACTPPEKAWMSI 377
Qy      61 LDRSOTRLQIVLISPELFIPVEDDVMERORLIESVDSVTPLIYEETDLMINIHDF 120
Db      378 LDRSOTRLQIVLISPELFIPVEDDVMERORLIESVDSVTPLIYEETDLMINIHDF 437
Qy      121 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIMEIA 180
Db      438 HVPFQSHHEEIEIFIFASECKTGRRLYKITSILKESKYRSSGGLPAPDFKCPIMEIA 497
Qy      181 ITSGEWEVLGRGNSIQVDEVRRLVFEETKOSPLEHLLYVSYVNPGEVTRLTDGYSH 240
Db      498 ITSGEWEVLGRGNSIQVDEVRRLVFEETKOSPLEHLLYVSYVNPGEVTRLTDGYSH 557
Qy      241 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 300
Db      558 SCCISOHCDFIFISKYSNOKNPHCVSLYKLSPEDDPTCKTKERWATILDSAGPLPYTPP 617
Qy      301 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQ----- 340
Db      618 EIFSFESTTGFTLYGMLYKPHDLQPKKYPVLFIYGGPQVOLLVNNRFKGVKXYRLNTLA 677
Qy      341 ----- 340
Db      678 SLGYYVVVVIDNRGSGRGLKFECAFKYKMGQIETDQVEGLQYLASRYPIDLDRYING 737
Qy      341 -----VALAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 380
Db      738 WSYGYLSLMLMQRSDIFRVAIAGAPVTLMIFYDTGYTERYMGHPDQNEQGYLGSVAM 797
Qy      381 QAEKFPSEBNRLLHLGFLDENVHFHTSILSLFVRACKPYDLQIYPERHSIRVPEG 440
Db      798 QAEKFPSEBNRLLHLGFLDENVHFHTSILSLFVRACKPYDLQIYPERHSIRVPEG 857
Qy      441 EHYELHLHYLOENLGSRIAALKVI 465
Db      858 EHYELHLHYLOENLGSRIAALKVI 882

RESULT 10

```

ABU07720
 ID ABU07720 standard; Protein; 882 AA.
 XX
 AC ABU07720;
 XX
 DT 19-MAY-2003 (first entry)
 XX
 DE Human serine protease HIPHUM46.
 XX
 KM Human; enzyme; HIPHUM46; serine protease; gene therapy; osteoarthritis;
 KM serine protease activity; modulation; dipeptidyl peptidase activity;
 KM musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
 KM Alzheimer's disease; paraneuronal palsy; Huntington's disease;
 KM amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
 KM irritable bowel syndrome; type 1 diabetes; faecal incontinence; anaemia;
 KM haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
 KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
 KM multiple sclerosis.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Region Location/Qualifiers
 FT 259..260
 FT /note= "Paired glutamates of the beta propeller domain"
 FT 739
 FT /label= Catalytic_serine_residue
 FT 817
 FT /label= Catalytic_aspartate_residue
 FT 849
 FT /label= Catalytic_histidine_residue
 FT Active-site
 XX
 PN GB2374869-A.
 XX
 PD 30-OCT-2002.
 XX
 PF 22-JAN-2002; 2002GB-0001404.
 XX
 PR 23-JAN-2001; 2001GB-0001760.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Edbrooke MR, Lewis AP;
 XX
 DR WPI; 2003-150703/15.
 DR N-PSDB; ABX12255.
 XX
 PT Identifying modulators of serine protease activity useful for treating
 PT musculoskeletal diseases, by contacting cell expressing a novel serine
 PT protease polypeptide with a compound and monitoring serine protease
 PT activity -
 XX
 PS Claim 10; Page 26-29; 38pp; English.
 XX
 CC The invention relates to a method of identifying a substance that
 CC modulates serine protease activity, comprising contacting a cell such as
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
 CC virus, expressing a serine protease polypeptide (HIPHUM 46), or its
 CC variant having dipeptidyl peptidase activity, or a serine protease
 CC isolated from the cell with a test substance and monitoring for serine
 CC protease activity. The method is useful for identifying a substance that
 CC modulates serine protease activity. A modulator of the serine protease that
 CC is useful in the manufacture of a medicament for treatment or prophylaxis of
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
 CC infection, Alzheimer's disease, paraneuronal palsy, myotonic
 CC dystrophy, Huntington's disease, or amyotrophic lateral sclerosis.
 CC Additional disease that may be treated using modulators of the serine
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung
 CC disease, type 1 diabetes, faecal incontinence, haemorrhoids, proctitis,
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
 CC sclerosis. The present sequence represents the amino acid sequence of the
 CC human serine protease HIPHUM46.
 XX

Sequence 882 AA;
 Query Match 0
 Best Local Similarity 97.6%; Score 2422; DB 24; Length 882;
 Best Local Similarity 82.3%; Pred. No. 1.6e-230;
 Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;
 QY 1 TGTANPKVTFKMSSEIMDAAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKYAWSI 60
 DB 318 TGTANPKVTFKMSSEIMDAAGRIIDVIDKELIQPFELFEGVEYIARAGTPEGKYAWSI 377
 QY 61 LDRSQRQLQIVLISELPIPVDDVMEQRQLIESVDSVTPLIYEETDWINHDF 120
 DB 378 LDRSQRQLQIVLISELPIPVDDVMEQRQLIESVDSVTPLIYEETDWINHDF 437
 QY 121 HVPQSHHEEIEFTFASSECKTGRHLKYTSLIKESKYRSSGGLPAPSDPKPIKEIA 180
 DB 438 HVPQSHHEEIEFTFASSECKTGRHLKYTSLIKESKYRSSGGLPAPSDPKPIKEIA 497
 QY 181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVYVNGEVTRLTDRGYSH 240
 DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLVYVYVNGEVTRLTDRGYSH 557
 QY 241 SCCTISQHCDFPISKYNNQKNPCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTP 300
 DB 558 SCCTISQHCDFPISKYNNQKNPCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTP 617
 QY 301 EIFSFESTGTFTLYGMLYKPHLOPKKXPTVLFITGGPO----- 340
 DB 618 EIFSFESTGTFTLYGMLYKPHLOPKKXPTVLFITGGPOVOLLNNRPFKGVKFRNLTLA 677
 QY 341 ----- 340
 DB 678 SLGYVVVVIDNRGSGRGLKFECAFYKMGQIEIDDOVEGLQYLASRYDFIDDRVGIGH 727
 QY 341 -----VAIAGNPLYTLMTFYDPTGYTERVYNGHDPDQEGYLLGSVAM 380
 DB 738 WSYGYLSLMAIMQSRDIFRVALAGAPVTLMTFYDPTGYTERVYNGHDPDQEGYLLGSVAM 797
 QY 381 QAEKFPSEPNRLLLLGFLDENVHFHTSILSLFLRAGKPYLQIYPOERHSIRVPSG 440
 DB 798 QAEKFPSEPNRLLLLGFLDENVHFHTSILSLFLRAGKPYLQIYPOERHSIRVPSG 857
 QY 441 EHYELHLHYLOENLGSRIALVKVI 465
 DB 858 EHYELHLHYLOENLGSRIALVKVI 882
 RESULT 11
 ID AAE14337
 ID AAE14337 standard; Protein; 580 AA.
 XX
 AC AAE14337;
 XX
 DT 07-MAR-2002 (first entry)
 XX
 DE Human protease PRS-2 protein.
 XX
 KM Human; protease; PRS-2; tranquilliser; gene therapy; vaccine; allergy;
 KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;
 KM atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;
 KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;
 KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;
 KM hypertension; neurological disorder; Parkinson's disease; drug screening;
 KM cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;
 KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;
 KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;
 KM developmental disorder; reproductive disorder; infertility; diarrhoea;
 KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice.
 XX
 OS Homo sapiens.
 XX
 PN W0200183775-A2.
 XX

PD 08-NOV-2001.
 XX 04-MAY-2001; 2001WO-US14651.
 PF 04-MAY-2000; 2000US-202082P.
 XX 11-MAY-2000; 2000US-203566P.
 PR 17-MAY-2000; 2000US-205803P.
 PR 25-MAY-2000; 2000US-207477P.
 PR 01-JUN-2000; 2000US-209402P.
 XX (INCYTE GENOMICS INC.
 PA Deleage AM, Lal P, Hafalia A, Patterson C, Wallia NK, Kearney L;
 PI Trippouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM;
 PI Reddy R, Yue H, Tang YT;
 XX WPI; 2002-034518/04.
 DR N-PSDB; AAD23843.
 XX Novel human proteases and polynucleotides encoding the proteases,
 PT useful for treating, diagnosing or preventing cell proliferative,
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental
 PT disorders -
 PS Claim 1; Page 120-121; 151pp; English.
 XX The invention relates to human proteases (PRTS-14) and its corresponding
 CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for
 CC the diagnosis, treatment and prevention of disorders associated with
 CC increased or decreased expression of PRTS. Examples of such disorders
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,
 CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders
 CC (AIDS, Addison's disease, allergy, anemia, asthma, atopic dermatitis,
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic
 CC infections); cardiovascular disorders (myocardial infarction, ischaemic
 CC heart disease and hypertension); neurological disorders (epilepsy,
 CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety
 CC and seasonal affective disorder and prion diseases); gastrointestinal
 CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder
 CC (infertility, disruption of estrous and menstrual cycle and
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).
 CC PRTS DNA is also in useful is gene therapy. PRTS and its immunogenic
 CC fragments are useful for screening libraries of compounds in several drug
 CC screening assays. The present sequence is human protease PRTS-2 protein.
 XX
 SQ Sequence 580 AA;
 Query Match 90.4%; Score 2244.5; DB 23; Length 580;
 Best Local Similarity 84.7%; Pred. No. 3.3e-213;
 Matches 436; Conservative 4; Mismatches 18; Indels 57; Gaps 4;
 QY 1 TGTANPKYTFKMSSEIMDAEGRIIDVIDELIOPFEILEGVEYIARAWTPEGKYANSI 60
 DB 73 TGTANPKYTFKMSSEIMDAEGRIIDVIDELIOPFEILEGVEYIARAWTPEGKYANSI 132
 QY 61 LLDROGTRQLQIVLISPELFIPEVDYMERQRLIESVPSVTLIYEETDITMINIHDF 120
 DB 133 LLDROGTRQLQIVLISPELFIPEVDYMERQRLIESVPSVTLIYEETDITMINIHDF 192
 QY 121 HVFPOSHEEIEIFIFASECKTGFRHLKYITTSILKSKYRSSGGLPAPS-DEKCPKIEBI 179
 DB 193 HVFPOSHEEIEIFIFASECKTGFRHLKYITTSILKSKYRSSGGLPAPVTMMITFMSL 252
 QY 180 AITSEMEVLGRHGSNIQVDEFRRLVYFEGCTDPSLEHLLVYVSVNPGCVTRLLDRGYS 239
 DB 253 GTPSC--MCVTHIVEIQVDEVRLRVYFEGTDSPLEHLLVYVSVNPGCVTRLLDRGYS 309

QY 240 HSOCISQHCDFPFIKSKYNOKNPHCVSLYKSSPEDDPCTCKTEFMAITLDSAGLPDYP 299
 DB 310 HSOCISQHCDFPFIKSKYNOKNPHCVSLYKSSPEDDPCTCKTEFMAITLDSAGLPDYP 369
 QY 300 PEIFSESTTGTFTLYGMLYKPHDLQPKKYPTVLFIYGGPO----- 340
 DB 370 PEIFSESTTGTFTLYGMLYKPHDLQPKKYPTVLFIYGGPOQLVNNRFRGVKYPRLNTL 429
 QY 341 -----VALAGPVTLMIFYDTGTYERTYKGHDPNE 370
 DB 430 ASIGYVVVVVIDNRGSCRLKEGAFKYMVVAIAGAPVTLMIFYDTGTYERTYKGHDPNE 489
 QY 371 QGYVYGSVMAQAEKPESEPNRLLHGFLENVHFANTSILSLFVRACKPYDLQIYPOE 430
 DB 490 QGYVYGSVMAQAEKPESEPNRLLHGFLENVHFANTSILSLFVRACKPYDLQIYPOE 545
 QY 431 RHSIRVPESEGEHYELHLHYLOENIGSRITAAKVI 465
 DB 546 RHSIRVPESEGEHYELHLHYLOENIGSRITAAKVI 580
 RESULT 12
 ABG61594
 ID ABG61594 standard; Protein; 690 AA.
 XX
 AC ABG61594;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-1 splice variant #1.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPP;
 KW DPPV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinnesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI; 2002-444178/47.
 DR N-PSDB; ABK83325.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PS Disclosure: Page 59-61; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPV)-related
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,

CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX Sequence 690 AA;

Query Match 72.9%; Score 1809.5; DB 23; Length 690;
 Best Local Similarity 94.0%; Pred. No. 6,2e-170;
 Matches 342; Conservative 6; Mismatches 3; Indels 13; Gaps 3;

QY 1 TGTANPKVTFKMEIMDAEGR11DIVDKELIQPEILFEGVEYIARAGMTPGKYAMS1 60
 DB 318 TGTANPKVTFKMEIMDAEGR11DIVDKELIQPEILFEGVEYIARAGMTPGKYAMS1 377
 QY 61 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 120
 DB 378 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 437
 QY 121 HVPQSHHEEIEFIFASECKTGRHLKYKTSILKESKYKRSGLPAPSDFCPIKEE1A 180
 DB 438 HVPQSHHEEIEFIFASECKTGRHLKYKTSILKESKYKRSGLPAPSDFCPIKEE1A 497
 QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFECTKDSPLHHLVYVSYNPGVTRLTDGYSH 240
 DB 498 ITSGEWEVLGRHGSNIQVDEVRLVYFECTKDSPLHHLVYVSYNPGVTRLTDGYSH 557
 QY 241 SCCISQHCDFITSKYSNQNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPP 300
 DB 558 SCCISQHCDFITSKYSNQNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPP 617
 QY 301 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFIYGG-----GPOVALIAGAPV----- 348
 DB 618 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFIYGRLLLLGPG-SLGGSSMIQDTRN 676
 QY 349 TLM1 352
 DB 677 VIMV 680

RESULT 13

ID ABG61600 standard; Protein; 658 AA.

AC ABG61600;

DT 12-AUG-2002 (first entry)

DE Human DPRP-1 splice variant #7.

XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.

OS Homo sapiens.

XX WO200231134-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.

XX 12-OCT-2000; 2000US-240117P.

XX (FERR) FERRING BV.

XX Qi S, Akinsanya KO, Riviere FJ, Juntien J;

DR MPI; 2002-444178/47.

XX N-PSDB; ABK83331.

DR

PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding

PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and

PT viral infections, cancers, allergies, neurological disorders, or pain

PT

XX

PS

XX Disclosure; Page 70-72; 113pp; English.

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 72.8%; Score 1808; DB 23; Length 658;

Best Local Similarity 100.0%; Pred. No. 8.1e-170; Indels 0; Gaps 0;

Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTANPKVTFKMEIMDAEGR11DIVDKELIQPEILFEGVEYIARAGMTPGKYAMS1 60
 DB 318 TGTANPKVTFKMEIMDAEGR11DIVDKELIQPEILFEGVEYIARAGMTPGKYAMS1 377
 QY 61 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 120
 DB 378 LDRSOTRLQIVLISPELFIPEVDVMERORLIESVPSVTPLIYEETTDIMINIHDI 437
 QY 121 HVPQSHHEEIEFIFASECKTGRHLKYKTSILKESKYKRSGLPAPSDFCPIKEE1A 180
 DB 438 HVPQSHHEEIEFIFASECKTGRHLKYKTSILKESKYKRSGLPAPSDFCPIKEE1A 497
 QY 181 ITSGEWEVLGRHGSNIQVDEVRLVYFECTKDSPLHHLVYVSYNPGVTRLTDGYSH 240
 DB 498 ITSGEWEVLGRHGSNIQVDEVRLVYFECTKDSPLHHLVYVSYNPGVTRLTDGYSH 557
 QY 241 SCCISQHCDFITSKYSNQNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPP 300
 DB 558 SCCISQHCDFITSKYSNQNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYTPP 617
 QY 301 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFIYGG 338
 DB 618 EIFSFESTTGFTLYGMLYKPHDLPQKKYPTVLFIYGG 655

RESULT 14

ID ABG61596 standard; Protein; 661 AA.

AC ABG61596;

DT 12-AUG-2002 (first entry)

DE Human DPRP-1 splice variant #3.

XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.

XX Homo sapiens.
 OS
 XX
 XX MO200231134-A2.
 XX
 XX
 XX 18-APR-2002.
 XX
 PF 12-OCT-2001, 2001MO-US31874.
 XX
 XX 12-OCT-2000, 2000US-240117P.
 XX
 XX (FERR) FERRING BV.
 XX
 XX Qi S, Akinsanya KO, Riviere PJ, Junien J;
 PI
 XX WPI, 2002-444178/47.
 DR N-PSDB; ABR83327.
 XX
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX
 XX Disclosure; Page 63-65; 113pp; English.
 XX
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
 XX
 XX Sequence 661 AA;
 SQ

Query Match 72.8%; Score 1808; DB 23; Length 661;
 Best Local Similarity 100.0%; Pred. No. 8, 1e-170;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGTANPKVTFFKSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 60
 DB 318 TGTANPKVTFFKSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 377
 QY 61 LLDROSTRLOIYLISPELFIPEVDVMEORLIESVPSVPLIYEETTDIWINIHDF 120
 DB 378 LLDROSTRLOIYLISPELFIPEVDVMEORLIESVPSVPLIYEETTDIWINIHDF 437
 QY 121 HVPQSHHEEIEFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 180
 DB 438 HVPQSHHEEIEFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 497
 QY 181 ITSGEWEVLGRHGSNIQVDEVRRLVFEGETKOSPLEHLLVYVYNGEVTRLTDGYSH 240
 DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVFEGETKOSPLEHLLVYVYNGEVTRLTDGYSH 557
 QY 241 SCCISGHCFFISKYSNOKNPHCVSLYKLSPPEDDCTCKTKEFWATILDSAGPLPYTTP 300
 DB 558 SCCISGHCFFISKYSNOKNPHCVSLYKLSPPEDDCTCKTKEFWATILDSAGPLPYTTP 617
 QY 301 EIFSESTGTFTLYGMLYKPHDLPQPKYPTVLFIYGG 338
 DB 618 EIFSESTGTFTLYGMLYKPHDLPQPKYPTVLFIYGG 655

RESULT 15
 ABG61601

ID ABG61601 standard; Protein; 613 AA.
 XX
 AC ABG61601;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPRP-1 splice variant #8.
 XX
 XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP,
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder.
 XX
 OS Homo sapiens.
 OS
 XX
 XX MO200231134-A2.
 XX
 XX 18-APR-2002.
 XX
 PF 12-OCT-2001, 2001MO-US31874.
 XX
 XX 12-OCT-2000, 2000US-240117P.
 XX
 XX (FERR) FERRING BV.
 XX
 XX Qi S, Akinsanya KO, Riviere PJ, Junien J;
 PI
 XX WPI, 2002-444178/47.
 DR N-PSDB; ABR83332.
 XX
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX
 XX Disclosure; Page 73-75; 113pp; English.
 XX
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
 XX
 XX Sequence 613 AA;
 SQ

Query Match 62.0%; Score 1540; DB 23; Length 613;
 Best Local Similarity 100.0%; Pred. No. 2, 8e-143;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGTANPKVTFFKSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 60
 DB 318 TGTANPKVTFFKSEIMIDABGR11DIVIKELIQPEILFEGVEYIARAGWTEGKYAWSI 377
 QY 61 LLDROSTRLOIYLISPELFIPEVDVMEORLIESVPSVPLIYEETTDIWINIHDF 120
 DB 378 LLDROSTRLOIYLISPELFIPEVDVMEORLIESVPSVPLIYEETTDIWINIHDF 437
 QY 121 HVPQSHHEEIEFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 180
 DB 438 HVPQSHHEEIEFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 497

QY 181 ITSGEWELGRHSNIQVDEVRLVYFEGTKDSPLEHLYVSVNPGVTRLTDRGYSH 240
Db 498 ITSGEWELGRHSNIQVDEVRLVYFEGTKDSPLEHLYVSVNPGVTRLTDRGYSH 557
QY 241 SCCISOHCDFFISKYSNOKNPHCVSLYKSSPEDDPTCKTKEFWATILDS 230
Db 558 SCCISOHCDFFISKYSNOKNPHCVSLYKSSPEDDPTCKTKEFWATILDS 607

Search completed: October 15, 2003, 17:10:46
Job time : 30.3565 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 10.3743 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-5

Perfect score: 2482

Sequence: 1 TGTANPKVTFKMSIEMDAE.....HLHYLOENLGSRIALAKVI 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6CTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfilltest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1377.5	55.5	310	4	US-09-794-236-4
2	309.5	12.5	771	4	US-09-462-284-2
3	238	9.6	766	4	US-10-002-593-6
4	234	9.4	755	5	PCT-US93-07923-3
5	234	9.4	759	5	PCT-US93-07923-2
6	234	9.4	766	1	US-08-230-491A-3
7	234	9.4	766	1	US-08-619-280A-3
8	234	9.4	766	2	US-08-940-391-3
9	234	9.4	766	4	US-09-794-236-1
10	214	8.6	760	1	US-08-230-491A-2
11	214	8.6	760	1	US-08-619-280A-2
12	214	8.6	760	2	US-08-940-391-2
13	139	5.6	632	3	US-09-016-080-1
14	106.5	4.3	593	3	PCT-US93-07923-11
15	102	4.1	732	3	US-08-989-299-5
16	102	4.1	1312	3	US-08-989-299-8
17	101	4.1	1313	3	US-08-989-299-9
18	97.5	3.9	782	4	US-09-585-858-29
19	94.5	3.8	2404	4	US-09-134-001C-3464
20	93	3.7	808	4	US-09-198-452A-327
21	92.5	3.7	721	4	US-09-390-234-20
22	92.5	3.7	1449	3	US-08-840-062-6
23	92.5	3.7	1723	4	US-09-194-612A-31
24	90.5	3.6	467	4	US-09-134-001C-3452
25	88	3.5	968	4	US-09-107-532A-5970
26	88	3.5	708	1	US-08-396-479B-8
27	88	3.5	708	1	US-08-818-823-8

28	88	3.5	739	1	US-08-396-479B-10	Sequence 10, Appl
29	88	3.5	739	1	US-08-818-823-10	Sequence 10, Appl
30	88	3.5	828	4	US-09-107-532A-4191	Sequence 4191, Ap
31	88	3.5	1068	1	US-08-396-479B-12	Sequence 12, Appl
32	88	3.5	1068	1	US-08-818-823-12	Sequence 12, Appl
33	88	3.5	1075	5	PCT-US94-07297-41	Sequence 41, Appl
34	87	3.5	1512	4	US-09-328-352-5163	Sequence 5163, Ap
35	86	3.5	657	4	US-09-355-166-1	Sequence 1, Appl1
36	86	3.5	821	1	US-07-935-311A-4	Sequence 4, Appl1
37	86	3.5	821	1	US-08-368-079-4	Sequence 4, Appl1
38	86	3.5	821	5	US-09-886-319A-33	Sequence 33, Appl1
39	86	3.5	821	5	PCT-US93-07996-4	Sequence 4, Appl1
40	86	3.5	1362	4	US-09-134-001C-4955	Sequence 4955, Ap
41	85.5	3.4	285	4	US-09-252-991A-27550	Sequence 27550, A
42	85.5	3.4	350	2	US-09-123-851-3	Sequence 3, Appl1
43	85.5	3.4	350	2	US-08-728-520-3	Sequence 3, Appl1
44	85.5	3.4	614	4	US-09-252-991A-20060	Sequence 20060, A
45	85	3.4	1193	4	US-09-227-725A-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81965/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-794-236-4

Query Match
Best Local Similarity 83.5%; Pred. No. 5,9e-142;
Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY	207	FEETDPLEHLLVYVSYVNPGEVTRLTDRGSHSCISOHCDPFIISKYSNKNPHCVSL	266
DB	1	FEETDPLEHLLVYVSYVNPGEVTRLTDRGSHSCISOHCDPFIISKYSNKNPHCVSL	60
QY	267	YKLSSPEDDPCTKTEFWATILDSAGPLPDYTPPEIFSEESTTGTLYGMLYKPHDLPQ	326
DB	61	YKLSSPEDDPCTKTEFWATILDSAGPLPDYTPPEIFSEESTTGTLYGMLYKPHDLPQ	120
QY	327	KKYPVLFITGGPO-----	340
DB	121	KKYPVLFITGGPOQOQIIDDQVEGLQYLASRYDFIDLDRAVGHGMSYGYLSLMAHQ	180
QY	341	-----VAIAGAPVTLMIFYDTGYTERVMGHPDQNGYLLSVAMQAKRPFSEPRLLLL	395
DB	181	SDIFVVAIAGAPVTLMIFYDTGYTERVMGHPDQNGYLLSVAMQAKRPFSEPRLLLL	240
QY	396	HGFLDENVHFATSTILSLFLVRAGKPYDLQIYQBRHSIRVPESGEHVELHLHYLOENL	455
DB	241	HGFLDENVHFATSTILSLFLVRAGKPYDLQIYQBRHSIRVPESGEHVELHLHYLOENL	300
QY	456	GSRIALAKVI 465	
DB	301	GSRIALAKVI 310	

RESULT 2

US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Michael
; APPLICANT: Van den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Fungus
US-09-462-284-2

Query Match 12.5%; Score 309.5; DB 4; Length 771;
Best Local Similarity 23.3%; Pred. No. 2.1e-24;
Matches 128; Conservative 67; Mismatches 174; Indels 181; Gaps 20;

QY 5 NPKVTFKMSSEIMIDAEGRILIDVIDKELIQPEILFEVGEYIARAGWTEGKYAMSILDR 64
DB 259 NPTVL-----LNTASKEVKAQPIDAESTDLI-----GEAW--LDT 298
QY 65 SOTRLQIVLISPELPIVEDDVMERORLIESVPSVTPLLIYEETDIMI-NIHDIHFV 123
DB 299 HTT-----VAAKAFNRVOD--QOKVAVDTASNKAT-VISDRDGTDLKDLRLSMKYT 349
QY 124 POSHEEIEIFIFASECKGFRHLKYITSLKESKXKRSSGGLPAPSDFCRPIKEIATIS 183
DB 350 PKPSDKAYVYIDISDHSGMAHLYLFP-----VSGGEPFIP-----LTK 387
QY 184 GEMEVLGHSNIIQVDEVRLVYFEGTKDPSLEHLLVYVSVNPGEVRLTDRGYS--HS 241
DB 368 GDMET-----SILSIDQERQLVYVLTSTGHSTERHLISYV-STFVAVPLVDVYAAKWS 442
QY 242 CCIISCHDFEISKYKNQKNPHCVSLYKSSPEDDPTCKTKEFWATILLASGFL---PDYT 298
DB 443 ASFSANSGVYILTYGGPDVPR-QELVYTTNS-----TKPL-RITIDNAKVLQGIKDYA 492
QY 299 PEIISFE--STGFTLYGMLYKPHDLQPGKKYPTVLFIYGP----- 339
DB 493 LPNITYFELPLPSGELTNVMORLPQFSPPKYPILLFTPYGGPGAQEVTKRMQALNPKAY 552
QY 340 ----- 339
DB 553 VASDESELEVYVTVNDRGTGFKGRKFRSAVTRQLGLEAEQIYAAQQAANIPIWDADHI 612
QY 340 -----QVALAGAPVTLMIFYDTGTTERYMGHPDNEGGYLLG 376
DB 613 GIWGSFGGYLTLSKYLEKDSGAFTLGVTAPYSDMRFPDSMTYTERYMKTLSTNEGEYETS 672
QY 377 SVAMQAEKPPSEPNRLLLHGFLENVHFAHTSILSLFVRAKGYDQIYPOEHSIRV 436
DB 673 AV-RKTDFKRVVEGFLIQHTGDDNVHFNQNSALVDLIMGD-----VSFEKLHSQWF 725
QY 437 PESGEHYELH 446
DB 726 TDSDHGISYH 735

RESULT 3
US-10-002-593-6
; Sequence 6, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:

; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTI
; TITLE OF INVENTION: INHIBITOR AND VASOPRESSIN INHIBITOR ASSOCIATED ANGIOEDEMA
; FILE REFERENCE: Acty Docket No. 6586198 1242/48/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patencin Version 3.1
; SEQ ID NO 6
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-593-6

Query Match 9.6%; Score 238; DB 4; Length 766;
Best Local Similarity 21.1%; Pred. No. 1.4e-16;
Matches 120; Conservative 71; Mismatches 193; Indels 186; Gaps 24;

QY 2 GNAPKVFKMSSEIMIDAEGRILIDVIDKELIQPEILFEVGEYIARAGWTEGKYAMSIL 61
DB 260 GAVNPTVAFV--VNTDSLSSVTNATSIQIAPASMLI-GDHYLCDVTMA----- 306
QY 62 LDRSOTRLQIVLISPELPIVEDDVMERORLIESVPSVTPLLIYEETDIMI-NIHDIHF 121
DB 307 ---QERISLQWL-----RIQNV--SVMDICQYDESSGMW-NC----- 339
QY 122 VFPQSHHEIEIFIFASECKGFRH-----LYKITSLSKSKYKRSSGGLPAPSDFCR 174
DB 340 LVARQHIEMSTTGWGRFPEPFTLDGNSFYKIIIS--NEEGYRHI-----CYFOID 390
QY 175 IKEEIAISGEMEVLGHSNIIQVDEVRLVYFEGT-KDSPLEHLLVYVSVNPGEVRL 233
DB 391 KKDCTFITKGTWEYIG-----IEALTSDLYTISNEYKMGFGKRNLYKIQLIDYKVTCL 445
QY 234 TDRGYSHSCIS-CHDFEISKYKNQKNPH--C-----VSLYKSSPEDDPTCKTKEFWA 285
DB 446 -----SCELNPERCQYVSFSKSAKYQLRCSGPELPYTLTHSSVNDGLRVLED-N 497
QY 286 TILDSAGLPPYTPPE--IFSESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGP-- 339
DB 498 SALDKM--LQNVQWPSKKLDFIILNETKF-WYOMILRPH-PDKSKYVPLLDVYAGPSCQ 553
QY 340 ----- 339
DB 554 KADIVFRLNMTYLAATENIIVASFDGRSGYQGDKIIMHAINRRLGTFEVEDQIEARQF 613
QY 340 -----QVALAGAPVTLMIFYDTGTTERYMG- 364
DB 614 SKMGFVDRKRIAIWGSYGVVTSVMVLSGSGVFKCGIAVAVPVRMEWYDVSYTERYVGL 673
QY 365 -HPQONEGGYLLSVAMQAEKPPSEPNRLLLHGFLENVHFAHTSILSLFVRAKGYD 423
DB 674 PTPEDNLDHYNSYTVMSAENF--KQVEYLLIHGTADNVHFNQNSALVDLIMGD 731
QY 424 LOIYPOEHSIRVPESEGEHYELHLYLOE 453
DB 732 AMWYTTDEDHGIASSTAQHITMHSPIKQ 761

RESULT 4
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schloesman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson

```

STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 755
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-3

Query Match          9.4%; Score 234; DB 5; Length 755;
Best Local Similarity 20.9%; Pred. No. 3,7e-16;
Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPVTFFKMSIMIDABGRRIIDVIDKELIOPELIFEGVEYIARAGTPECKYAMSL 61
DB 249 GAVNPTVKFFV--VNTDSLSSVTNATSIQTAPASMLI-GDHYLCOVTA----- 295
QY 62 LDRSQRLQIVLISPELIFVEDDVMERORLIESVPSVTPLIYEETDWINIHDIHF 121
DB 296 ---TGRISLQWL-----RIGNY--SVMDICDYDESSGRW-NC----- 328
QY 122 VFPQSHBEIEFIIFASECKTGFRH-----LYKITSILKESYKXSSGGLPAPSDPKCP 174
DB 329 LVARQHIEIMSTGWGRFRPSEPHFTLDGNSFYKII--NEEGYRHI-----CYFOID 379
QY 175 IKEIAlTSGEMEVUGRHGNSIQVDEVRRLVFEGR-KDSPLHHLVYVSVNPGEVTRL 233
DB 380 KDCCTFITKGTWEVIG-----IEALTSIDLYYISNEYKMPGGRNLYKI-----OL 425
QY 234 TDRGYSHSCIS-----QHCDFFISKYSNQNKP-----C-----VSLYKLSPEDDPTCKTK 281
DB 426 SD--YRKVTCLSGELNPERQYVSFSKAKYQLRCSGPGPLVTLIHSVNDKGLRVL 483
QY 282 EFNATILDSAGPLPDYTPPE--IFSESTTGTLYGMLYKPPHDLPGKKYPTLVFIYGG 338
DB 484 ED-NSALDKM--LQNVQMPSKLDFIILNETKF-WYQMLPRH-PDKSKYVPLLDVAVG 538
QY 339 P-----R-----R-----R-----R-----R-----R-----R----- 339
DB 539 PCSQKADTVFRLMWTATYLASTENIYASFDGSGSYQGDKIMAINRRLGTFEVEDQIEA 598
QY 340 -----QVALAGAVTLMIIFYDTGTYER 361
DB 599 ARQFSKGFVDNKRIRAIWGMVSYGYVTSMTVLGSGVFKGGLAIVAVSNMEYDSYTYER 658
QY 362 YMG--HPDQNBQCYIYGVSVAQAEKPESEPNRLLLHGLFDENVHFAHYSILISFLVRAG 419

```

```

DB 659 YMGLEPTPEDNLDRYRNSVTWSRAENF--KQVEYLLHGTADNVHFOQSAQISKALVDVG 716
QY 420 KPYDLQIYQERHSIRVPSGGEHYELHLYLOE 453
DB 717 VDFQAMWYDEDDGIASTAHQHIYTHMSHFIKQ 750

RESULT 5
PCT-US93-07923-2
Sequence 2, Application PC/TUS9307923
GENERAL INFORMATION:
APPLICANT: Morimoto, Chikao
APPLICANT: Schloesman, Stuart F.
APPLICANT: Tanaka, Toshiaki
TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 759
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-2

Query Match          9.4%; Score 234; DB 5; Length 759;
Best Local Similarity 20.9%; Pred. No. 3,7e-16;
Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPVTFFKMSIMIDABGRRIIDVIDKELIOPELIFEGVEYIARAGTPECKYAMSL 61
DB 253 GAVNPTVKFFV--VNTDSLSSVTNATSIQTAPASMLI-GDHYLCOVTA----- 299
QY 62 LDRSQRLQIVLISPELIFVEDDVMERORLIESVPSVTPLIYEETDWINIHDIHF 121
DB 300 ---TGRISLQWL-----RIGNY--SVMDICDYDESSGRW-NC----- 332
QY 122 VFPQSHBEIEFIIFASECKTGFRH-----LYKITSILKESYKXSSGGLPAPSDPKCP 174
DB 333 LVARQHIEIMSTGWGRFRPSEPHFTLDGNSFYKII--NEEGYRHI-----CYFOID 383
QY 175 IKEIAlTSGEMEVUGRHGNSIQVDEVRRLVFEGR-KDSPLHHLVYVSVNPGEVTRL 233
DB 384 KDCCTFITKGTWEVIG-----IEALTSIDLYYISNEYKMPGGRNLYKI-----OL 429

```



```

Oy      234 TDRYSHCISCI-----OHQDFISIKYNOXKPH---C---VSLYKLSDEPDPTCTK 281
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      430 SD--YTKVTLSCGLNPERCOYYSVSEFSAKAYQOLRGSGFRLPYLTHSVNDKGRVL 487
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      282 EFMAWTLDSAGPLPDYTPPE--ISFSESTGTLGYMLYKPHDLOPSKKYPTVLFYGG 335
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      488 ED-NSALDKM--LONNOMPSKKLDLFIILNETKF-WYOMILPEH-FDKSKKYPPLLDVYAG 544
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      339 P----- 335
          |
Dh      543 PCSOKADYVRLMAYTLASTENIIVASFDRGSGGYGDKIMHINRLGTFVEVDQIEA 602
          |
Oy      340 -----OVIAGAPVLMIFDYDTCTYER 367
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      603 AROFSKMGFVDNKRRIAMGWSYGYVTSMVLSGSGSVFPCGIALAPVSRMEYUOSVYTER 662
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      362 YMG--HPDOWEGGYLLGSVAMQAEKFPSEBNRLLLHGFDEBNVFAFTHSILSLFVYAG 419
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      663 YMGLPTEPDNLDHYRNVSTWRAENF--KQVEYLLIHGTRADNVHFGQSAQISKALVDVG 722
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      420 KPYDLOIYPOERHSIRVPESGEHYELHLHLHYLOE 453
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      721 VDFQAMMYTDEDDGIASTAHQIITYHMSHFIIQ 754
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

1      RESULT 6
2      US-08-230-491A-3
3      : Sequence 3, Application us/08230491A
4      : Patent No. 55872299
5      : GENERAL INFORMATION:
6      :   APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
7      :   APPLICANT: Gartin-Chessa, Pilar; Old, Lloyd J.
8      :   TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
9      :   TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
10     :   TITLE OF INVENTION: THEREOF
11     :   NUMBER OF SEQUENCES: 10
12     :   CORRESPONDENCE ADDRESS:
13     :   ADDRESSEE: FELFE & LYNCH
14     :   STREET:
15     :   CITY: NEW YORK
16     :   STATE: NEW YORK
17     :   COUNTRY: USA
18     :   ZIP: 10022
19     :   COMPUTER READABLE FORM:
20     :   MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
21     :   COMPUTER: IBM PS/2
22     :   OPERATING SYSTEM: PC-DOS
23     :   SOFTWARE: WORDPERFECT - ASC II
24     :   CURRENT APPLICATION DATA:
25     :   APPLICATION NUMBER: US/08/230.491A
26     :   FILING DATE: 20-APRIL-1994
27     :   CLASSIFICATION: 435
28     :   ATTORNEY/AGENT INFORMATION:
29     :   NAME: Hanson, No. 55872299man D.
30     :   REGISTRATION NUMBER: 30,946
31     :   REFERENCE/DOCKET NUMBER: LUD 330
32     :   TELECOMMUNICATION INFORMATION:
33     :   TELEPHONE: (212) 688-9200
34     :   TELEFAX: (212) 838-3884
35     :   INFORMATION FOR SEQ ID NO: 3:
36     :   SEQUENCE CHARACTERISTICS:
37     :     LENGTH: 766 amino acids
38     :     TYPE: amino acid
39     :     TOPOLOGY: linear
40     :   US-08-230-491A-3
41
42     Query Match          9.4%; Score 234; DB 1; Length 766;
43     Best Local Similarity 20.9%; Pred. No. 3.8e-16;
44     Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;
45
46     2  GTANKKVFKNKSEIMIDAGRIIDVIDKELIIPFELLEGVEYIARAGMTPEGKYAMSL 61
47     260  GAVNITVTFEV--VNFDSLSVTNATSTIOITPAASMLT-GDHYLCDDVTVA----- 306

```

```

OY 62 LBSORLQOVLISPELFI PVEDVMERQOLIESPDSVPFLIEETDQWIMHIFH 123
Db 307 ---TQERISQWL-----RIQNT--SYMDICDYDESSGN-NC----- 335
OY 122 VPDOSHEEIEFI FASECTGRH-----LYKITSILKESKYRSQGLPABDFKCP 174
Db 340 LVARQHIEIMSTGVMVGRFRSEPHFTLQDONSFYKITS--NEGYRHI-----CYFOID 390
OY 175 IKEIATITGSEWVLVGRHSGNIOVDEVARLYFEET--KDSPLEHNLVYVSVNPGVTRL 233
Db 391 KKQCTITTKGTMEVIO-----LEALTSDLYIISNEYKMGPOGRKRLYKI-----QL 433
OY 234 TDRGYSHSCIS-----QHCDFFISKYSNOKNPH--C-----VSLYKLSPEDDPTCKTK 281
Db 437 SD--YTKVTCLSCELPNERCQYVSFSFEAKYVQLRCSGQPLRYLTLSVNDKGLRVL 494
OY 282 EFPAITLDSAGRLPDVTPE---ISFESTGFTLYGLYKLRPHDQREKKYPTVLYYCG 338
Db 495 ED--NSALDKM--LQNVQMPSSKLDFTILNETKF--WYQMI LPPH--FDKSKYLPRLLDVYAG 549
OY 339 P----- 339
Db 550 PCSQKADTVFLMWATYLASTENIIVASFDRGSGYQDKIMHAINRLGTFEEVDQIEA 608
OY 340 -----QVAIACAPRTLMFVDTQYTER 361
Db 610 AROFSKMGFVNDKRIAIWGSVGYVTSMTVLGSGGVFKGJIAVAPVSRMEYDVSYTER 668
OY 362 YMG--HPDNOEGQYLYGSVAMQAEKFPSEPNRLLLHGFLEDNVHFATSTILSFLVRAG 419
Db 670 YMGILPREDNLHDYRNSVTWSRAENF--KQVEYLLIHGTADDNVHFQGSQADISKALVDVG 722
OY 420 KPYDQLIYPOERHSIRVRESGSHYELHLNHYLOE 453
Db 728 VDFQAMMYTDEDDGJASSTAHQHIYTHMSHFTKO 761

```

RESULT 7
 US-08-619-280A-3
 Sequence 3, Application US/08619280A
 Patent No. 5767242
 GENERAL INFORMATION:
 APPLICANT: Zimmermann, Rainer; Park, John E.;
 APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
 TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
 TITLE OF INVENTION: ALPHA, AND USES THEREOF
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Felle & Lynch
 STREET: 805 Third Avenue
 CITY: New York City
 STATE: New York
 COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: PC-DOS
 SOFTWARE: Wordperfect
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/619,280A
 FILING DATE: 18-MARCH-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/230,491
 FILING DATE: 20-APRIL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Hanson, No. 5767242man D.
 REGISTRATION NUMBER: 30,946
 REFERENCE/DOCKET NUMBER: LUD 5330.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 766 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-619-280A-3

Query Match

9.4%; Score 234; DB 1; Length 766;
 Best Local Similarity 20.9%; Pred. No. 3.8e-16;

Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPKTTFKMSIIMDAEGRIDVIDKELIOPELIFEGVEYIARAGWTPCKYASIL 61
 DB 260 GAVNPYTKFV--VNTDSLSTVNATSIQTAPASMLI-GDHYLCVDTWA----- 306
 QY 62 LDRSQRLQIVLISPELIFPVEDVMERORLIESVDSVTPLLIYEETDWINIHDI 121
 DB 307 ---TOERISLOWL-----RRIQNY--SVMDICDYDESSGRW-NC----- 339
 QY 122 VFPQSHHEEIEFIPASECKTGFRH-----LYKITSILKESYKXSSGGLPAPSDPKP 174
 DB 340 LVARQHIEMSTGTVGGRFPRSEPHFTLDGNSFYKILS--NEEGYRHI-----CYFOID 390
 QY 175 IKEEIAITSGEWEVLGRHGSNIQVDEVRRLVFEPT-KDSPLBHLVYVSVNPGEVTRL 233
 DB 391 KKOCFTITKGTWEVIG-----IEALTSDYLYISNEYKMGPGGRNLTKI-----QL 436
 QY 234 TDRGYSHSCCIS-----QHCDFFISKYSNQNPH---C---VSLYKLSPEDDPTCKTK 281
 DB 437 SD--YTKVTCLSCELNPERCOYVSFSKAKYQLRCSGPGLPYTLHSSVNDKGLRVL 494
 QY 282 EFWATILDSAGPLPDYTPPE---IFSFEPTGFTLYGMLYKPHLDQPKKYPTVLYFG 338
 DB 495 ED-NSALDKM--LQNVQMPKSLDFIILNETKF-WYOMILPPH-FDKSKKYPLLDDVYAG 549
 QY 339 P-----RRIQNY--SVMDICDYDESSGRW-NC----- 339
 DB 550 PCSQKADTVFRLNMTATLASTENITIVASPDGRSGYQGDKIMHAINRRLGFEVEDOIEA 609
 QY 340 -----OVALAGAPVTLMIFYDTGYTER 361
 DB 610 AROFSKMGFVNDKRIAIWGSYGYVTSMVLGSGGVFKCIGIAPVPSRMEYDVSYYTER 669
 QY 362 YMG--HPDQNEQGYLGSAVMAQAEKPPSEPNRLLLHGFLDEVNHFATSHLSFLVRAG 419
 DB 670 YMGLEPTPEDNLDHRYNSTVMSRAENF--KOVEYLLHGTADNVHFQOQAISKALVDVG 727
 QY 420 KPYDLQIYPOERHSIRVPESGEHYELHLHYLOE 453
 DB 728 VDFQAMWYTDDEHGIASTAHQHLYTHMSHPIKQ 761

RESULT 8

US-08-940-391-3

; Sequence 3, Application US/08940391

; Patent No. 5965373

; GENERAL INFORMATION:

; APPLICANT: Zimmermann, Rainer; Park, John E.;

; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.

; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION

; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felle & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage

; COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS
 SOFTWARE: Wordperfect
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/940,391
 FILING DATE: 01-OCT-1997
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/619,280
 FILING DATE: 18-MARCH-1996
 APPLICATION NUMBER: 08/230,491
 FILING DATE: 20-APRIL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Hanson, No. 5965373man D.
 REGISTRATION NUMBER: 30,946
 REFERENCE/DOCKET NUMBER: LUD 5330.1
 TELEPHONE: (212) 688-9200
 TELECOMMUNICATION INFORMATION:
 TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 766 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-940-391-3

Query Match

9.4%; Score 234; DB 2; Length 766;
 Best Local Similarity 20.9%; Pred. No. 3.8e-16;

Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

QY 2 GTANPKTTFKMSIIMDAEGRIDVIDKELIOPELIFEGVEYIARAGWTPCKYASIL 61
 DB 260 GAVNPYTKFV--VNTDSLSTVNATSIQTAPASMLI-GDHYLCVDTWA----- 306
 QY 62 LDRSQRLQIVLISPELIFPVEDVMERORLIESVDSVTPLLIYEETDWINIHDI 121
 DB 307 ---TOERISLOWL-----RRIQNY--SVMDICDYDESSGRW-NC----- 339
 QY 122 VFPQSHHEEIEFIPASECKTGFRH-----LYKITSILKESYKXSSGGLPAPSDPKP 174
 DB 340 LVARQHIEMSTGTVGGRFPRSEPHFTLDGNSFYKILS--NEEGYRHI-----CYFOID 390
 QY 175 IKEEIAITSGEWEVLGRHGSNIQVDEVRRLVFEPT-KDSPLBHLVYVSVNPGEVTRL 233
 DB 391 KKOCFTITKGTWEVIG-----IEALTSDYLYISNEYKMGPGGRNLTKI-----QL 436
 QY 234 TDRGYSHSCCIS-----QHCDFFISKYSNQNPH---C---VSLYKLSPEDDPTCKTK 281
 DB 437 SD--YTKVTCLSCELNPERCOYVSFSKAKYQLRCSGPGLPYTLHSSVNDKGLRVL 494
 QY 282 EFWATILDSAGPLPDYTPPE---IFSFEPTGFTLYGMLYKPHLDQPKKYPTVLYFG 338
 DB 495 ED-NSALDKM--LQNVQMPKSLDFIILNETKF-WYOMILPPH-FDKSKKYPLLDDVYAG 549
 QY 339 P-----RRIQNY--SVMDICDYDESSGRW-NC----- 339
 DB 550 PCSQKADTVFRLNMTATLASTENITIVASPDGRSGYQGDKIMHAINRRLGFEVEDOIEA 609
 QY 340 -----OVALAGAPVTLMIFYDTGYTER 361
 DB 610 AROFSKMGFVNDKRIAIWGSYGYVTSMVLGSGGVFKCIGIAPVPSRMEYDVSYYTER 669
 QY 362 YMG--HPDQNEQGYLGSAVMAQAEKPPSEPNRLLLHGFLDEVNHFATSHLSFLVRAG 419
 DB 670 YMGLEPTPEDNLDHRYNSTVMSRAENF--KOVEYLLHGTADNVHFQOQAISKALVDVG 727
 QY 420 KPYDLQIYPOERHSIRVPESGEHYELHLHYLOE 453
 DB 728 VDFQAMWYTDDEHGIASTAHQHLYTHMSHPIKQ 761

RESULT 9

US-09-794-236-1

; Sequence 1, Application US/09794236

```

; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-1

```

```

Query Match          9.4%; Score 234; DB 4; Length 766;
Best Local Similarity 20.9%; Pred. No. 3.8e-16;
Matches 120; Conservative 70; Mismatches 190; Indels 194; Gaps 25;

```

```

QY      2 GTANPKVTFKMSLIMDAEGRHIDVIDKELIQPEILEGVEYIARAGWTEGKAWSL 61
DB      260 GAVNFTVAFV--VNTDLSSTYNTATSIQITAPASMLI-GDHYLCDVTMA----- 306
QY      62 LDRSQTRLQIVLISELFIPEVDYMERORLIESVPSVTPLLIYEETDIIWINDIFH 121
DB      307 ---TQERISLQWL-----RRIQNY--SYMIDICVDDESSGRW-NC----- 339
QY      122 VFPOSHHEIEFIIFASECKTGFRH-----LYKTSILSKSKYRSSGGLPAPSDPKP 174
DB      340 LVAQOHIEMTSTTGWGRFRPEPHFTLDGNSPFYKII--NEEGKPHI-----CYFOID 390
QY      175 IKEEIAITSGEMVELGRHGSNIQVDEVARLVYFEGT-KDSPLEHLLVYVSVNPEGEVRL 233
DB      391 KKDCTFITKGTMEVIG-----IEALTSPLYIISNEHYKMGGRLLYKI-----QL 436
QY      234 TDRGYSHSCIS----OHCDFIISKYSNQKNPH--C-----VSLYKLSPEDDPTCKT 281
DB      437 SD--YTKVTCISCEINPERCOYVSFSKKAQYOLRCSGPLPLYTLHSSVNDKGLRVL 494
QY      282 EFMAITLDSAGRLPDYTPRE---IFSFESTTGFTLYGMLYKPHDQPKKYPVFIYGG 338
DB      495 ED-NSALDKM--LQNVQMPKSLDFTIINETKF-WYQMLPPH-FDKSKKYPPLLDVYAG 549
QY      339 P----- 339
DB      550 PCSQKADTVFRILNMTYLASTENIIVASFDRGSGCYGQDKIMHAINRLGTFEVEDQIEA 609
QY      340 -----OVAIAGAPVTLMIFDGTGTER 361
DB      610 AROFSKMGFVNDKRIAIWGSYGYVTSMVLGSGGVFCGIAVAPVSRMEYDSVYTER 669
QY      362 YMG--HPDONGGYVLGSVAMQAEKFPSEPNRLILLHGFLENHFAFATSIILSLVYAG 419
DB      670 YMLGPTPEPDNDLHDYNSITVMSAENP--KQVEYLIIHGADNVAFQOASQISKALVDVG 727
QY      420 KPYDLQIYPOERHSIRVPESEGEHYELHLHYLOE 453
DB      728 VDFQAMWYTDDEDHGIASTAHQHIYTHMSHFIKQ 761

```

```

RESULT 10
US-08-230-491A-2
; Sequence 2, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN, AND USES
; NUMBER OF SEQUENCES: 10

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FELLE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-230-491A-2

```

```

Query Match          8.6%; Score 214; DB 1; Length 760;
Best Local Similarity 20.2%; Pred. No. 5.8e-14;
Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;

```

```

QY      140 KTGFRHLKYKTSILSKSKYRSSGGLPAPSDPKPIKEEIAITSGEMVELGRHGSNIQVD 199
DB      371 KDGKTHIHYI-----KQVENAIQIITSGKMEAI-----NI--- 400
QY      200 EVRLVYFEGTKDSPLEHLLVYVSVNPEGEVRLTDRGYSHS-CCISQH-----CDFEIS 253
DB      404 -----F-FRYTQDSLFFYSNMFEEYPPGRNRIYRISISGYPSSKKCVTCHLAKERCQYYTA 453
QY      254 KYSNQKNPHCVSLY----KLSPEDDPTCKTKFMAITLDSAGRLPD-----YTPPEIFS 304
DB      454 SFSYAKYVYALVCYGPGLPISTLHDGRTDGE---IKILEENKELNALKNIQLPKEEIK 509
QY      305 FESFTTGFTLYGMLKPHDLOQPKKYPVLFYGGP----- 339
DB      510 KLEVDKITLMYKMLLPQFPDRSKKYPILLIQYGGPSCQSVRSVAVNMWISTLASKEGVI 569
QY      340 ----- 339
DB      570 ALVDGRTAFOGDKLLVAVYRKLGVEVEVDQITAVRKFIEMGFIDEKRIAIWGSYGYV 629
QY      340 -----OVAIAGAPVTLMIFDGTGTERYMGHP--DQNEGGYVLGSVAMQAEKF 385
DB      630 SSLAIASGTGLFKCGIIVAPVSVWEYVASVYTERFMGLPTKDDNLHFKYSTVVARAEYF 689
QY      386 PSEPNRLILLHGFLENHFAFATSIILSLFVLRACKPYDLQIYPOERHSIRVPEG---EH 442
DB      690 RNVD--YLLIHGTADDNVHFQNSAQIAKALVNAQVDQAMYSIDQNHGL-----SGLSTNH 743
QY      443 YEHLHLHYLOE 453
DB      744 LYTHMTHFLKQ 754

```

```

RESULT 11
US-08-619-280A-2
; Sequence 2, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;

```

APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
NUMBER OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-619-280A-2

Query Match
Best Local Similarity 20.2%; Score 214; DB 1; Length 760;
Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;

QY 140 KTFPHLKYITSLKSKYKSSGGLPAPDFKCPKEIAITSGEWELGHRGSIQVD 199
DB 371 KDGYKHIIYI-----KDTVENAIOITSGKWEAI-----NI----- 400

QY 200 EVRRLVPEGTGKDSPLBHLVYVYNPGEVTRLDRGYSHS-CCISQH-----CDFIS 253
DB 401 -----FRVTDLSLFYSSNEFEETPGRRNIYRISIGSYPSKCCVTLKREKCOQYTA 453

QY 254 KYSNOKNHCVSLEY-----KLSSPEDDPTCKTEFWATILDSAGPLPD-----YTPPEIFS 304
DB 454 SFSDYAKYVYALVCGPGIPISLHDGRDQE-----IKILENKELMNLKNIQLPKREIK 509

QY 305 FESITGFLYGLMLYRPHLDQPKKPYTVLIFYGCP----- 339
DB 510 KLEVEITLWYKMLLPQFDRSKYRPLIOYVGGPCOSVRSVFAVNMISYLASKEGMVI 569

QY 340 ----- 339
DB 570 ALVDGRGTAPOGDKLLVAVYRKLGYEVEDQITAVRKFIEMGFIDEKRIAIWGSYGGYV 629

QY 340 -----OVAIAGAPVTLMIFYDTGYTERYMGHP--DQNEOGYVLGSVMAQAEKF 385
DB 630 SSLALASGTGLFKCGIAVAVPSSWEYVASYTERFMGLPTKODNLEHYKNSTVMAAREYF 689

QY 386 PSEPRKLLILGFLDENVHFATSTILSLVLNACRPYDIOIYPOERHSIRVEGSG--EH 442
DB 630 RNVD--YLLIHGTADNVHFOHSAQIAKALVNAOVDFOAMWYSQDNHGL-----SGLSITH 743

QY 443 YELHLLHYLOE 453
DB 744 LYTHMTHFLKO 754

RESULT 12
US-08-940-391-2
Sequence 2, Application US/08940391
Patent No. 5965373
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280
FILING DATE: 18-MARCH-1996
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5965373man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-940-391-2

Query Match
Best Local Similarity 20.2%; Score 214; DB 2; Length 760;
Matches 87; Conservative 53; Mismatches 127; Indels 164; Gaps 13;

QY 140 KTFPHLKYITSLKSKYKSSGGLPAPDFKCPKEIAITSGEWELGHRGSIQVD 199
DB 371 KDGYKHIIYI-----KDTVENAIOITSGKWEAI-----NI----- 400

QY 200 EVRRLVPEGTGKDSPLBHLVYVYNPGEVTRLDRGYSHS-CCISQH-----CDFIS 253
DB 401 -----FRVTDLSLFYSSNEFEETPGRRNIYRISIGSYPSKCCVTLKREKCOQYTA 453

QY 254 KYSNOKNHCVSLEY-----KLSSPEDDPTCKTEFWATILDSAGPLPD-----YTPPEIFS 304
DB 454 SFSDYAKYVYALVCGPGIPISLHDGRDQE-----IKILENKELMNLKNIQLPKREIK 509

QY 305 FESITGFLYGLMLYRPHLDQPKKPYTVLIFYGCP----- 339
DB 510 KLEVEITLWYKMLLPQFDRSKYRPLIOYVGGPCOSVRSVFAVNMISYLASKEGMVI 569

QY 340 ----- 339
DB 570 ALVDGRGTAPOGDKLLVAVYRKLGYEVEDQITAVRKFIEMGFIDEKRIAIWGSYGGYV 629

QY 340 -----OVAIAGAPVTLMIFYDTGYTERYMGHP--DQNEOGYVLGSVMAQAEKF 385
DB 630 SSLALASGTGLFKCGIAVAVPSSWEYVASYTERFMGLPTKODNLEHYKNSTVMAAREYF 689

[illegible][illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 : Search time 65.704 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-5

Perfect score: 2482
Sequence: 1 TGTANPKVTFKMSIMIDAE.....HLHYLGSLSRALAKVI 465

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2422	97.6	882	10 US-09-976-674-1	Sequence 1, Appli
2	2422	97.6	882	12 US-10-054-776-2	Sequence 2, Appli
3	2422	97.6	882	12 US-10-170-789-38	Sequence 38, Appli
4	1809.5	72.9	690	10 US-09-976-674-7	Sequence 7, Appli
5	1808	72.8	658	10 US-09-976-674-19	Sequence 19, Appli
6	1808	72.8	613	10 US-09-976-674-11	Sequence 11, Appli
7	1540	62.0	663	10 US-09-976-674-21	Sequence 21, Appli
8	1494.5	60.2	863	10 US-09-976-674-3	Sequence 3, Appli
9	1494.5	60.2	892	10 US-09-976-674-23	Sequence 23, Appli
10	1494.5	60.2	892	10 US-09-976-674-27	Sequence 27, Appli
11	1445	58.2	879	10 US-09-976-674-33	Sequence 33, Appli
12	1445	58.2	879	10 US-09-976-674-35	Sequence 35, Appli
13	1377.5	55.5	310	12 US-09-993-959-4	Sequence 4, Appli
14	1187.5	47.8	508	15 US-10-237-271-3	Sequence 3, Appli
15	1094.5	44.1	832	10 US-09-976-674-29	Sequence 29, Appli

16	1094.5	44.1	832	10 US-09-976-674-31	Sequence 31, Appli
17	1045	42.1	819	10 US-09-976-674-37	Sequence 37, Appli
18	1045	42.1	819	10 US-09-976-674-39	Sequence 39, Appli
19	471.5	19.0	518	10 US-09-976-674-25	Sequence 25, Appli
20	321	12.9	129	11 US-09-764-891-3564	Sequence 3564, Ap
21	277	11.2	818	12 US-10-401-437-3	Sequence 3, Appli
22	277	11.2	818	12 US-10-402-312-3	Sequence 3, Appli
23	277	11.2	818	12 US-10-402-067-3	Sequence 3, Appli
24	277	11.2	818	12 US-10-401-436-3	Sequence 3, Appli
25	267	10.8	93	9 US-09-867-550-988	Sequence 988, Ap
26	249.5	10.1	767	15 US-10-165-603-4	Sequence 4, Appli
27	247	10.0	706	10 US-09-976-674-11	Sequence 41, Appli
28	247	10.0	796	10 US-09-976-674-5	Sequence 5, Appli
29	247	10.0	796	12 US-09-870-133-2	Sequence 2, Appli
30	247	10.0	796	15 US-10-160-501-5	Sequence 5, Appli
31	238	9.6	766	12 US-10-423-714-6	Sequence 6, Appli
32	238	9.6	766	14 US-10-002-593-6	Sequence 6, Appli
33	238	9.6	766	15 US-10-165-603-7	Sequence 7, Appli
34	234	9.4	766	9 US-09-265-606-3	Sequence 3, Appli
35	234	9.4	766	12 US-09-993-959-1	Sequence 1, Appli
36	214	8.6	760	9 US-09-265-606-2	Sequence 2, Appli
37	214	8.6	760	12 US-10-301-822-55	Sequence 55, Appli
38	214	8.6	760	15 US-10-177-293-136	Sequence 136, App
39	214	8.6	803	12 US-10-205-219-58	Sequence 58, Appli
40	211.5	8.5	691	10 US-09-976-674-43	Sequence 43, Appli
41	190.5	7.7	710	15 US-10-156-761-10681	Sequence 10681, A
42	165	6.6	164	15 US-10-106-598-6740	Sequence 6740, Ap
43	146	5.9	432	15 US-10-237-271-4	Sequence 4, Appli
44	112	4.5	358	10 US-09-976-674-13	Sequence 13, Appli
45	104	4.2	1436	9 US-09-815-242-5566	Sequence 5566, Ap

ALIGNMENTS

RESULT 1
US-09-976-674-1
Sequence 1, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akimsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976.674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240.117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 882
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-674-1

Query Match 97.6% Score 2422; DB 10; Length 882;
Best Local Similarity 82.3% Pred. No. 1.4e-231;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;
QY 1 TGTANPKVTFKMSIMIDAEGRRIIDVDKELIOPELIFEGVEYIARAGMTPEGKYAMSI 60
DB 318 TGTANPKVTFKMSIMIDAEGRRIIDVDKELIOPELIFEGVEYIARAGMTPEGKYAMSI 377
QY 61 LLDSSQRLQIVLSPFLFVEDDVMERGLISSVPSVTPLIYETTDIWINIHIF 120
DB 378 LLDSSQRLQIVLSPFLFVEDDVMERGLISSVPSVTPLIYETTDIWINIHIF 437
QY 121 HVPFQSHHEEIEFTFASCECKTGFRHLKYTSLKESKYKRSGLPAPSDPCPIKEBIA 180
DB 438 HVPFQSHHEEIEFTFASCECKTGFRHLKYTSLKESKYKRSGLPAPSDPCPIKEBIA 497

Oy	181	ITSGEMEVLAGRGNSIQVDEVARLVYFEGTKDSPLEHLVYVSVNPGCVTRLTDGYSH	244
Db	498	ITSGEMEVLAGRGNSIQVDEVARLVYFEGTKDSPLEHLVYVSVNPGCVTRLTDGYSH	557
Oy	241	SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPCTCKTEFMATILDSAGLPDYTP	300
Db	558	SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPCTCKTEFMATILDSAGLPDYTP	617
Oy	301	EIPFSESTGFLYCMLYKPHDLPCKKPYTLFLYIGSPQ-----	340
Db	618	EIPFSESTGFLYCMLYKPHDLPCKKPYTLFLYIGSPQVOLVNNRFQVKYFRNLTLA	677
Oy	341	-----	340
Db	678	SLGYVVVVIADNNGSCHRLKFECAFYKMKQJIEIDDPQVEGLQYLASRYDPIDLDVRGIGH	737
Oy	341	-----VNIAGAPTYLMTFYDNGYNERVMGHDPDQNGQGYLGSVAM	380
Db	738	WSYGYLSIMALMQSDIFRVAIAGAPVTLMTFYDNGYNERVMGHDPDQNGQGYLGSVAM	797
Oy	381	QAEKPESEBNRLLLHGFLENVVPHNTSILSLFLVRACKPYDLQIYPERHSIRVPESG	440
Db	798	QAEKPESEBNRLLLHGFLENVVPHNTSILSLFLVRACKPYDLQIYPERHSIRVPESG	857
Oy	441	EHYELHLHYDQENIGSRIALAKVI	465
Db	858	EHYELHLHYDQENIGSRIALAKVI	882

RESULT 2
US-10-054-776-2

```

: Sequence 2, Application US/10054776
: Publication No. US20030165818A1
: GENERAL INFORMATION:
: APPLICANT: Mark Robert Edbrooke
: APPLICANT: Alan Peter Lewis
: TITLE OF INVENTION: NOVEL PROTEIN
: FILE REFERENCE: OQ1042US
: CURRENT APPLICATION NUMBER: US/10/054, 776
: CURRENT FILING DATE: 2002-01-23
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 2
: LENGTH: 882
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-054-776-2

```

Query Match	97.6%	Score 2422	DB 12	Length 882
Best Local Similarity	82.3%	Pred. No. 1,4e-231		
Matches	465	Conservative	0	Mismatches 0
			Indels	100, Gaps 1
OY	1	TGTANPKYTKFKSEIMIDAEGRRIIDYIDKELLQPFILREGVEYIARACMTBEGKAWSI	60	
Db	318	TGTANPKYTKFKSEIMIDAEGRRIIDVIDELLQPFILREGVEYIARACMTBEGKAWSI	377	
OY	61	LIDRSQTRLQIYLISPELLPIVEDDVMERQRIIESVPDSVTPLIIEEFTDMINIHDF	120	
Db	378	LIDRSQTRLQIYLISPELLPIVEDDVMERQRIIESVPDSVTPLIIEEFTDMINIHDF	437	
OY	121	HVFPSHSEELIFFITASECKTGFRRHLYKTTSLIKESKYRSSGGLPABEDFKCIKEELA	180	
Db	438	HVFPSHSEELIFFITASECKTGFRRHLYKTTSLIKESKYRSSGGLPABEDFKCIKEELA	497	
OY	181	ITSGEMEVLGRHGSNIQVDEVARLYVFEGTKDPSLEHLHYVVSYYNPGSVTRLDRGYSH	240	
Db	498	ITSGEMEVLGRHGSNIQVDEVARLYVFEGTKDPSLEHLHYVVSYYNPGSVTRLDRGYSH	557	
OY	241	SCCIIHQHDFPISKSNQKNPHCVSLYKLSPEDDPCTCKEFMAITLDSAGPLDYTPP	300	
Db	558	SCCIIHQHDFPISKSNQKNPHCVSLYKLSPEDDPCTCKEFMAITLDSAGPLDYTPP	617	

OY	301	EIFSESTTGFLTYMLTKRPHDLOGKKXPVLYFYGGPQ-----	340
Dd	618	EIFSESTTGGFLTYGMLKRPHDLOGKKIPVLYFYGGPQVLNNRFKGVKFRNLTLA	677
OY	341	-----	340
Dd	678	SLGYVVVIIDNRGSCHRGKFEGAEFKYMKQIIEDDQVEGLQYLASRYFDIDLDRVGING	737
OY	341	-----VAIAAPPTLWTFYDTGYTERFMGHDPDNOEGYYLSGVAM	380
Dd	738	WSYGYSLLMALMQRSDIFRAVIAAPPTLMFTPDGYTERFMGHDPDNOEGYYLSGVAM	797
OY	381	QAEPFSPENRLLLLHGFLENVHPAHTSILLSPVRACKPYDLQIYPQERHSIRVPESG	440
Dd	798	QAEPFSPENRLLLLHGFDEVHPAHTSILLSPLVRACKPYDLQIYPQERHSIRVPESG	857
OY	441	EHYEHLHLHYLOENIGSRITAAIKVI	465
Dd	858	EHYEHLHLHYLOENIGSRITAAIKVII	882

RESULT 3
US-10-170-789-38
Sequence 38, Application US/10170789
Publication No. US20030180930A1
GENERAL INFORMATION:
APPLICANT: Rachel E. Meyers
APPLICANT: Olandt, Peter J.
APPLICANT: Kapeller-Liberman, Rosana
APPLICANT: Curtis, Rory A. J.
APPLICANT: Willemson, Mark
APPLICANT: Welch, Nadine
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF
FILE REFERENCE: 10448-191001
CURRENT APPLICATION NUMBER: US/10/170,789
CURRENT FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US 09/797,039
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/06525
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/186,061
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 09/882,166
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: PCT/US01/19269
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/212,078
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 09/934,406
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: PCT/US01/26052
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: US 60/226,740
PRIOR FILING DATE: 2000-08-21
PRIOR APPLICATION NUMBER: US 09/861,801
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: PCT/US01/16549
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 60/205,508
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: US 09/801,267
PRIOR FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: PCT/US01/07138
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 60/197,454
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/829,671
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: PCT/US01/40463
PRIOR FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: US 60/197,508
PRIOR FILING DATE: 2000-04-18

```

; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-789-38
```

```

Query Match      97.6%; Score 2422; DB 12; Length 882;
Best Local Similarity 82.3%; Pred. No. 1,4e-231;
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;
```

```

QY 1 TGTANPKYTFKMSSEIMDAEGRIIDVIDKELIQPELIFEGVEYIARAGWTEGKYANSI 60
DB TGTANPKYTFKMSSEIMDAEGRIIDVIDKELIQPELIFEGVEYIARAGWTEGKYANSI 377
QY 61 LLDRSQTRLQIVLISPELFIPVEDDVMERORLIESVPSVPLIYEETTDIMINIDIF 120
DB LLDRSQTRLQIVLISPELFIPVEDDVMERORLIESVPSVPLIYEETTDIMINIDIF 437
QY 121 HVPFQSHHEEIEFIFASCEKTFGRHLKYITSILKSKYKSSGGLPAPSDFCPIKEEIA 180
DB HVPFQSHHEEIEFIFASCEKTFGRHLKYITSILKSKYKSSGGLPAPSDFCPIKEEIA 497
QY 181 ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLYVVSYNVPEVTRLTDRGYSH 240
DB ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLYVVSYNVPEVTRLTDRGYSH 557
QY 241 SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 300
DB SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 617
QY 301 EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPO-----GFOVAIAGAPV----- 348
DB EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPO-----GFOVAIAGAPV----- 677
QY 341 -----VAIAGAPVTLWTFYVYTGTERYMGPPDQNEGYIIGSIAM 380
DB -----VAIAGAPVTLWTFYVYTGTERYMGPPDQNEGYIIGSIAM 737
QY 341 -----VAIAGAPVTLWTFYVYTGTERYMGPPDQNEGYIIGSIAM 380
DB -----VAIAGAPVTLWTFYVYTGTERYMGPPDQNEGYIIGSIAM 737
QY 381 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFVRAGKPYDQIQOERHSIRVPESSG 440
DB 381 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFVRAGKPYDQIQOERHSIRVPESSG 440
QY 798 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFVRAGKPYDQIQOERHSIRVPESSG 857
DB 798 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFVRAGKPYDQIQOERHSIRVPESSG 857
QY 441 EHYELHLHYLOENUGSRIALAKVI 465
DB 441 EHYELHLHYLOENUGSRIALAKVI 465
QY 858 EHYELHLHYLOENUGSRIALAKVI 882
DB 858 EHYELHLHYLOENUGSRIALAKVI 882
```

```

RESULT 4
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-7
```

```

Query Match      72.9%; Score 1809.5; DB 10; Length 690;
Best Local Similarity 94.0%; Pred. No. 8.4e-171;
Matches 342; Conservative 6; Mismatches 3; Indels 13; Gaps 3;
```

```

QY 1 TGTANPKYTFKMSSEIMDAEGRIIDVIDKELIQPELIFEGVEYIARAGWTEGKYANSI 60
DB TGTANPKYTFKMSSEIMDAEGRIIDVIDKELIQPELIFEGVEYIARAGWTEGKYANSI 377
QY 61 LLDRSQTRLQIVLISPELFIPVEDDVMERORLIESVPSVPLIYEETTDIMINIDIF 120
DB LLDRSQTRLQIVLISPELFIPVEDDVMERORLIESVPSVPLIYEETTDIMINIDIF 437
QY 121 HVPFQSHHEEIEFIFASCEKTFGRHLKYITSILKSKYKSSGGLPAPSDFCPIKEEIA 180
DB HVPFQSHHEEIEFIFASCEKTFGRHLKYITSILKSKYKSSGGLPAPSDFCPIKEEIA 497
QY 181 ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLYVVSYNVPEVTRLTDRGYSH 240
DB ITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLYVVSYNVPEVTRLTDRGYSH 557
QY 241 SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 300
DB SCGISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFMATIIDSAGPLDPYTP 617
QY 301 EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPO-----GFOVAIAGAPV----- 348
DB EIFSESTTGFTLYGMLYKPHDLPQKKYPTVLFYGGPO-----GFOVAIAGAPV----- 676
QY 349 TLWI 352
DB 349 TLWI 352
DB 677 VIWV 680
```

```

RESULT 5
US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-976-674-19

Query Match 72.8%; Score 1808; DB 10; Length 658;

Best Local Similarity 100.0%; Pred. No. 1,1e-170;

Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 TGTANPKVTFKXSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 60
DB 318 TGTANPKVTFKXSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 377
OY 61 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 120
DB 378 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 437
OY 121 HVPFOSHEEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 180
DB 438 HVPFOSHEEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 497
OY 181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDRGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDRGYSH 557
OY 241 SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWATILDSAGPLPDYTPP 300
DB 558 SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWATILDSAGPLPDYTPP 617
OY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 338
DB 618 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 655
```

RESULT 6

US-09-976-674-11

Sequence 11, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akimsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.1

SEQ ID NO 11

LENGTH: 661

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-674-11

Query Match 72.8%; Score 1808; DB 10; Length 661;

Best Local Similarity 100.0%; Pred. No. 1,1e-170;

Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 TGTANPKVTFKXSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 60
DB 318 TGTANPKVTFKXSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 377
OY 61 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 120
DB 378 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 437
OY 121 HVPFOSHEEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 180
DB 438 HVPFOSHEEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 497
OY 181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDRGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDRGYSH 557
```

```
OY 241 SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWATILDSAGPLPDYTPP 300
DB 558 SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWATILDSAGPLPDYTPP 617
OY 301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 338
DB 618 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYIGG 655
```

RESULT 7

US-09-976-674-21

Sequence 21, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akimsanya, Karen

APPLICANT: Riviere, Pierre

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.1

SEQ ID NO 21

LENGTH: 613

TYPE: PRT

ORGANISM: Homo sapiens

US-09-976-674-21

Query Match 62.0%; Score 1540; DB 10; Length 613;

Best Local Similarity 100.0%; Pred. No. 4,6e-144;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 TGTANPKVTFKXSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 60
DB 318 TGTANPKVTFKXSEIMIDAGRIIDVIDKELIQPEILFEGVEYIARAGMTPEGKYAWSI 377
OY 61 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 120
DB 378 LIDRSQTRIQIVLISPELFIPEDDVMEORLIESVPSVTPLIYEETTDIWINIHDI 437
OY 121 HVPFOSHEEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 180
DB 438 HVPFOSHEEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFCPIKEEIA 497
OY 181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDRGYSH 240
DB 498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHLLYVVSYPNPGVTRLTDRGYSH 557
OY 241 SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWATILDS 290
DB 558 SCCISQHCDFPISKYSNQNPHCVSLYKLSPEDDPTCKTKEFWATILDS 607
```

RESULT 8

US-09-976-674-3

Sequence 3, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Qi, Steve

APPLICANT: Akimsanya, Karen

APPLICANT: Riviere, Pierre

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

CURRENT FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 863
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-674-3

Query Match 60.2%; Score 1494.5; DB 10; Length 863;
Best Local Similarity 50.5%; Pred. No. 2.7e-139;
Matches 281; Conservative 69; Mismatches 105; Indels 101; Gaps 2;

QY 1 TGTANPKYTFKMSLIMDAEGRILIDVDEKLIQPELLEFEGVEYIARAGWTPEGKYANSI 60
DB TGSKNPKIALKLAETQTSQCKIVSTQEKELVQFPSSLPFKVEYIARAGWTGKYANAM 367
QY 61 LIDRSQTRQLVILISPELFIVEDDVMERQRLIESVPSVTLIIYEETDWINIHDF 120
DB FLDRPQOMLOLVLPALFIPSTENEBORLASARAVPRNVQPVYVEETVWIMVHDF 427
QY 121 HVFPQSH-EEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDKCPIKEI 179
DB YFPQSEBDELCFLRANECKTGFCHLYKVAVLSQGYDWSPPSPGEDKCPIKEI 487
QY 180 AITSGEWEVLGRHGSNIQDEVRLVYEGTKDPSLEHLLVYVSVNGEVTRLTDGYS 239
DB ALTSGEWEVLARHGSKIWNBEETKLVPFGTKDTPLEHLLVYVYEAGEIVRLTTGFS 547
QY 240 HSCCISQHCDFPISKYSNQNKPCHVSLYKSSPEDDPTCKTKEFWATILDSAGPLDYTP 299
DB HSCSMQNFDMFVSHYSSVSTPCVHVYKLSGPDDELPHKQPRFMAWMMAASCPDYVP 607
QY 300 PEIFSESTGTGTLGMLYKPHDLOPKKYPVTLFIYGGPQ----- 340
DB PEIFHFHRSQVRLYGMILYKPHALQPKGHPVTLFVYGPOVLVNSFKIKYLRNLTL 667
QY 341 ----- 340
DB ASLGYAVVVIDGRSCQGRLEFEGALKNQMGQVEIEDQVEGLQFAEKYGFIDLSRVAIH 727
QY 341 ----- 340
DB 668 ASLGYAVVVIDGRSCQGRLEFEGALKNQMGQVEIEDQVEGLQFAEKYGFIDLSRVAIH 727
QY 341 ----- 340
DB 728 GMSYGGFLSLMGLIHKPQVFKVAILAGAPVTWMAADTYTERYKMDVPENNQGYAGSVA 787
QY 380 MQAEKPSPEPNRLILHGFIDENVAFAHTSILSLFLVAGKRYDQIYPOEHSIRVES 439
DB LHVETLPNEPNRLILHGFIDENVAFAHTSILSLFLVAGKRYDQIYPOEHSIRVES 847
QY 440 GEHYELHLHYLOENL 455
DB 848 GEHYEVTLLHFLQEYL 863

RESULT 9
US-09-976-674-23
Sequence 23, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976.674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240.117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patentin version 3.1
SEQ ID NO 23
LENGTH: 892
TYPE: PRT

ORGANISM: Homo sapiens
US-09-976-674-23

Query Match 60.2%; Score 1494.5; DB 10; Length 892;
Best Local Similarity 50.5%; Pred. No. 2.7e-139;
Matches 281; Conservative 69; Mismatches 105; Indels 101; Gaps 2;

QY 1 TGTANPKYTFKMSLIMDAEGRILIDVDEKLIQPELLEFEGVEYIARAGWTPEGKYANSI 60
DB TGSKNPKIALKLAETQTSQCKIVSTQEKELVQFPSSLPFKVEYIARAGWTGKYANAM 396
QY 61 LIDRSQTRQLVILISPELFIVEDDVMERQRLIESVPSVTLIIYEETDWINIHDF 120
DB FLDRPQOMLOLVLPALFIPSTENEBORLASARAVPRNVQPVYVEETVWIMVHDF 456
QY 121 HVFPQSH-EEIEFIASECKTGFRLHYKITSILKESKYKSSGGLPAPSDKCPIKEI 179
DB YFPQSEBDELCFLRANECKTGFCHLYKVAVLSQGYDWSPPSPGEDKCPIKEI 516
QY 180 AITSGEWEVLGRHGSNIQDEVRLVYEGTKDPSLEHLLVYVSVNGEVTRLTDGYS 239
DB ALTSGEWEVLARHGSKIWNBEETKLVPFGTKDTPLEHLLVYVYEAGEIVRLTTGFS 576
QY 240 HSCCISQHCDFPISKYSNQNKPCHVSLYKSSPEDDPTCKTKEFWATILDSAGPLDYTP 299
DB HSCSMQNFDMFVSHYSSVSTPCVHVYKLSGPDDELPHKQPRFMAWMMAASCPDYVP 636
QY 300 PEIFSESTGTGTLGMLYKPHDLOPKKYPVTLFIYGGPQ----- 340
DB PEIFHFHRSQVRLYGMILYKPHALQPKGHPVTLFVYGPOVLVNSFKIKYLRNLTL 696
QY 341 ----- 340
DB 697 ASLGYAVVVIDGRSCQGRLEFEGALKNQMGQVEIEDQVEGLQFAEKYGFIDLSRVAIH 756
QY 341 ----- 340
DB 757 GMSYGGFLSLMGLIHKPQVFKVAILAGAPVTWMAADTYTERYKMDVPENNQGYAGSVA 816
QY 380 MQAEKPSPEPNRLILHGFIDENVAFAHTSILSLFLVAGKRYDQIYPOEHSIRVES 439
DB LHVETLPNEPNRLILHGFIDENVAFAHTSILSLFLVAGKRYDQIYPOEHSIRVES 876
QY 440 GEHYELHLHYLOENL 455
DB 877 GEHYEVTLLHFLQEYL 892

RESULT 10
US-09-976-674-27
Sequence 27, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPEIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976.674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240.117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patentin version 3.1
SEQ ID NO 27
LENGTH: 892
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-674-27
Query Match 60.2%; Score 1494.5; DB 10; Length 892;
Best Local Similarity 50.5%; Pred. No. 2.7e-139;

	Matches	281;	Conservative	69;	Mismatches	105;	Indels	101;	Gaps	2
Oy	1	TGTANPKYTFKMSSEIMDAEGRIIDVIDKELOIPFELFEQVEYIARAGMTPEBGKYANSI	60							
Dd	337	TGSKNPKALKLAEEQTODSOGKIVASTOEKELQPSSSLFPKKVEYIARAGMTDGRYAAM	396							
Oy	61	LDRSGTRLQVILISPELFIPIEDDMVERORLIIESVPDSVTPLIYEETTDIMINIHDF	120							
Dd	397	FUDRQOVLQVLLPFLPALFISTENBEQRALABARAVPRVQCYYVVEEVTNWVINHDF	456							
Oy	121	HVFPOSH-EEIEIFIASECKTGPHLYKITYSILKESKYKRSSGGLPADSDFCPIKEEI	179							
Dd	457	YPFPOSEEDELCIFFRANECKTGFCMLKYVTAVLKSOGDWSEBSPESPGEDEFKCEPIKEEI	516							
Oy	180	AITSGEMEVLAGHSNSIQVDEVRALIYPEGTSDPLEHLVVYSVYNBSEVRLTRDRGS	239							
Dd	517	ALTSGEMEVLAGHSKIVNESETKLVPFGTVKOTPEHLHYVSYTEABAGEIYRLTLTPGFS	576							
Oy	240	HSCCISOHCDFEITSIKYSNOKNPHCVSLYLQSSPEDPTCKTKEFNATILDSAGLPDYPP	299							
Dd	577	HSCSNQSQNDMVSHTSYSSVSTPCVCHVYKLSGPDDDP.LHKQCRFNASMWMAASCPDVP	636							
Oy	300	PEIFSFEFTGGTYLGMLYKPHDLQPGKKYPIVYLFYGGPO-----	340							
Dd	637	PEIFFHRTSRDVRLKGMITYKPHALOFGKKHPVLVYGGPOQLVNNSFKGIKYLRLNTL	696							
Oy	341	-----	340							
Dd	697	ASLGAUVVVIIDRGSGCQGRLEFGALXKNMGQVEIEDQVEJGLQFAEKGYFDLSRVANH	756							
Oy	341	-----VALGAPVTLMIFFDTGTERMYGHDPONDECGYLYGSVA	379							
Dd	757	GWSYGFLSLMGLIHKKPVFKVAILAGAPTVMMAADTGTERMYMVDPENNQHGYENGSAVA	816							
Oy	380	MOAEKFSEPENRLLHLHGFLDENVHPAHRISILSFVLRAGKRYDDQIYPOEHSISRVPES	439							
Dd	817	LHVELLPDENRLLHLHGFLDENNVHFPHTNFLVSOLIRAKGYOLOIYPNERHSIRCPBS	876							
Oy	440	GEHYELHLLHYLOENL 455								
Dd	877	GEHYEVTLLHFLQEYL 892								
 RESULT 11 US-09-976-674-33 ; Sequence 33, Application US/09976674 ; Patent No. US20020115843A1 ; GENERAL INFORMATION: ; APPLICANT: Qi, Steve ; APPLICANT: Akimsanya, Karen ; APPLICANT: Riviere, Pierre ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV ; FILE REFERENCE: 70669 ; CURRENT APPLICATION NUMBER: US/09/976,674 ; CURRENT FILING DATE: 2001-10-12 ; PRIOR APPLICATION NUMBER: US 60/240,117 ; PRIOR FILING DATE: 2000-10-12 ; NUMBER OF SEQ ID NOS: 61 ; SOFTWARE: PatentIn version 3.1 ; SEQ ID NO 33 ; LENGTH: 879 ; TYPE: PRT ; ORGANISM: Homo sapiens US-09-976-674-33										
Oy	Query Match	58.2%; Score 1445; DB 10; Length 879;								
	Best Local Similarity	49.3%; Pred. No. 2,3e-134;								
Dd	Matches 274; Conservative	67; Mismatches 101; Indels 114; Gaps 3								
Oy	1 TGTANPKYTFKMSSEIMDAEGRIIDVIDKELOIPFELFEQVEYIARAGMTPEBGKYANSI	60								
Dd	337 TGSKNPKALKLAEEQTODSOGKIVASTOEKELQPSSSLFPKKVEYIARAGMTDGRYAAM	396								

QY	61	LLDSSQRLQVLVLSPELFIIVEDDWMERGLLISVPSVPLIIYEETDIIWIMHIF	120
DB	399	FLDRPQMLQVLVLPALFPISTENEBOQLASARAVPNVQYVVEVTWIMVNHDF	456
QY	121	HVFPOSH-ESEIEFIASECKTGRHLKYKTSILKESKYKSSGGLPAPSPFKCPKEI	179
DB	457	YFPQGBEEDLCLFRANECKTGCHLYKVAVLKSGQYDMSPEPSPEDEKCIKEI	516
QY	180	AITGSEVLCRHSNIOVDEVRLVYFEGTKDSFLEHLLYVSVYVNGEYTRLIDRGYS	239
DB	517	ALTGSEWVLLARHGS-----KGTKDPLEHLLYVSYEAGAEIYRLTTPGFS	563
QY	240	HSCCISQCHDFFISKYSNQKNPHCVSLYKLSSPEDDPCKTKEPATIILDSAGLPDYTP	299
DB	564	HSCNSQNFDMFVSHSYSSVSTPCCVHYVKSJGPDDDLHKRPFRWASMEALSCPPDYV	623
QY	300	PEIFSFEFTTFTLYGMLYKPHDLQPKKYPTVLFTYGGPO-----	340
DB	624	PEIFHFHRSVRLXGMLYKXHALQPKKHFTVLFTYGGPQVQVYVNSFKGIKYLRLNTL	683
QY	341	-----	340
DB	684	ASLGAVVVINDRSGSCQGLRFEQALKNQMGVEIEDQVEGLQFAEKYGFIDLSRVAIH	743
QY	341	-----VATAGAPVTLMIPDPCYTERMYMHPRONQEGCYLGSVA	379
DB	744	GWSYGGFLSLMGLIHKPOVFYKVALAGAVTYWMAVDITYTERMYMVPENNQGYAGSVA	803
QY	380	MOAEKFPSEPNRLLLHGFLENDENVFAHTSILLSFLVAGKRPYDQIYPOEKHSIRVES	439
DB	804	LHVELLPNEPRRLILHGFLENDENVFPTNTFLVSQLIRAGKRPYQIYIPNEHHSIRCPES	863
QY	440	GEHYELHLHYLOENL 455	
DB	864	GEHYEVTLLHFLQEYL 879	
RESULT 12			
US-09-976-674-35			
/ Sequence 35, Application US/09976674			
/ Patent No.0US20020115843A1			
/ GENERAL INFORMATION:			
/ APPLICANT: Qi, Steve			
/ APPLICANT: Akinsanya, Karen			
/ APPLICANT: Riviere, Pierre			
/ APPLICANT: Junien, Jean-Louis			
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV			
/ FILE REFERENCE: 70669			
/ CURRENT APPLICATION NUMBER: US/09/976,674			
/ PRIOR FILING DATE: 2001-10-12			
/ PRIOR APPLICATION NUMBER: US 60/240,117			
/ NUMBER OF SEQ ID NOS: 61			
/ SOFTWARE: PatentIn version 3.1			
/ SEQ ID NO 35			
/ LENGTH: 879			
/ TYPE: PRT			
/ ORGANISM: Homo sapiens			
US-09-976-674-35			
Query/ Match			
Best Local Similarity 49.3%; Score 1445; DB 10; Length 879;			
Matches 274; Conservative 67; Mismatches 101; Indels 114; Gaps 3			
QY	1	TGTANPKYTFKMSIIMIDABGRIIDVIDKELIQPELILFEGVEYIARAGWTEGKYANSI	60
DB	337	TGSNNPKIALKLAEPQDTSQCKIVSTQEKELVQPPSSLFPRKVEYIARAGWTRDKGYAAM	396
QY	61	LLDSSQRLQVLVLSPELFIIVEDDWMERGLLISVPSVPLIIYEETDIIWIMHIF	120
DB	397	FLDRPQMLQVLVLPALFPISTENEBOQLASARAVPNVQYVVEVTWIMVNHDF	456

QY	121	HYFPQSH-EEETLEFIPASCKTGFRHLYTITSLTKSKYKRSSGGU.PAPSDFCPIKEE	179
Db	457	YFFPOSGBEDELCTFLANECKTGFCRLKYATAVKSGQIDWSEPSFGDEDFCPIKEE	516
QY	180	AITSGMEVLAGRHGSIQIDEVARLTVFEGTKDPSLEHHLYVVSYNVPGCVTLTTRGYS	239
Db	517	ALTSGEVETLARGHGS-----KGTOTPLEHHLYVVSYEAAAGSIVALTTPGFS	563
QY	240	HSCCISQHCDFPISKYSNOKNPHCVSLYKILSSPEDDPTCKTKFPMATILDSAGPLPDYTP	299
Db	564	HSCSSONQNDMVSVSHSYSTPCVYVYKLSGDDDBLHPORPFMA5MMEAACPDYVP	623
QY	300	PEIFSPESSTGTGLYQMLTKYPHDLOQKKYPTLYFYGGPQ-----	340
Db	624	PEIFHFHTSDVRLYIMITKPHALQFQKHPVLYFYGGPQVOLLVNN5FGICYRLRLNTL	663
QY	341	-----	340
Db	684	ASLGAVVVYIDGRGSCQRLRFEGALKNQMGVEIEDQVGLQFVAKGYFDILSRVAIH	743
QY	341	-----VAIAGAPLTIMFYPTDQTERFMGHPDQNEQGYILGSVA	379
Db	744	GWSYGGFLSLMGLIHKPQVFKVAIAAPATVMAAYDTGTERMDVBPENNQHGYEAGSVA	803
QY	380	MOAEKFPSEPNRLILLHGFLDENVHAHTSILSLFVRAKGKPYDLOIYPOERHSIRVPES	439
Db	804	LHVEKLTPNPNRLILLHGFLDENVHFFHNLFLVLSQILRBGKPYQLQIYPERHSIRCPES	863
QY	440	GEHYELHLHYIOENL	455
Db	864	GEHYEVTLHLFLOEYL	879

QY	396	HGFLENNHFAHTSLLSFLVAGKDPYDLOIYPOERHSTRVPESGGEYHLLHLYAOENL	455
Db	211	HGFLENNHFAHTSLLSFLVAGKRPYDLOIYPOERHSTRVPESGGEYHLLHLYAOENL	300
QY	456	GSRIALRYI	465
Db	301	GSRIALRYI	310

```

RESULT 14
US-10-237-271-3
; Sequence 3, Application US/10237271
; Publication No. US20030096328A1
; GENERAL INFORMATION:
; APPLICANT: THE BURNHAM INSTITUTE
; APPLICANT: SMITH, Jeffrey W.
; APPLICANT: KRIDEL, Steven J.
; APPLICANT: AXELROD, Fumiko T.
; TITLE OF INVENTION: SRINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS
; FILE REFERENCE: BURNI100-1
; CURRENT APPLICATION NUMBER: US/10/0237,271
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 60/317,842
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-237-271-3

```

```

RESULT 13
US-09-993-959-4
; Sequence 4, Application US/09993959
; Publication No. US20030165489A1
; GENERAL INFORMATION:
; APPLICANT: Grouzman, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/993, 959
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-959-4

Query Match      55.5%; Score 1377.5; DB 12; Length 310;
Best Local Similarity 83.5%; Pred. No. 2,3e-128;
Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1

Qy      207  FEGTDSPLEHLLVYVSYVNPGEVYRLTRRGVSHSCCISOHQDFFISKYSNQNPHCVSL 266
Db      1   FEGTDSPLEHLLVYVSYVNPGEVYRLTRRGVSHSCCISOHQDFFISKYSNQNPHCVSL 60

Qy      267  YKLSPEDDPTCKTEKFMAITIDSAGPLPDYTPPELIFSPESITGFTLYGMLYKPHDLP 326
Db      61  YKLSPEDDPTCKTEKFMAITIDSAGPLPDYTPPELIFSPESITGFTLYGMLYKPHDLP 120

Qy      327  KKYPTVLEIYGGPO----- 340
Db      121  KKYPTVLEIYGGPOCOIEIDQVEGHQYLASRYDFIDLRVGIHGSYGYLSLMAIMQR 180

Qy      341  -----VALNAGPVYTWIPIFDYGTGYTERYKMGHPONNEGYYLGSVANOAEKFPSEPNRL 395
Db      181  SDIFVALNAGPVYTWIPIFDYGTGYTERYKMGHPONNEGYYLGSVANOAEKFPSEPNRL 240

```

[illegible]

Db 506 EYL 508

```

RESULT 15
US-09-976-674-29
; Sequence 29, Application US/0997674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-29
    
```

```

Query Match      44.1%; Score 1094.5; DB 10; Length 832;
Best Local Similarity 57.3%; Pred. No. 1.5e-99;
Matches 197; Conservative 58; Mismatches 88; Indels 1; Gaps 1;

QY      1 TGTANPKVTFKMSGIMIDAGRIIDVIDKELIQFELLFEGVEYIARAGWTPEGKYAMS1 60
DB      337 TGSKNPKRIAKLAEFOTDSQGIIVSTQEKELVQFSSLFPKVEYIARAGWTRDGKYAMAM 396
QY      61 LDRSQTRLQIVLISPELPIVEDDWMERQLIESVDSVTPLIIEETDININHDIF 120
DB      397 FLDRPOQWLOLVLLPPLALFIPSTENEOQLASARAAPRNQPYVVEEVTNWVNHDI 456
QY      121 HVPFOSH-EEIEFIFASECKTGFRLHLYTISILKESKYRSSGGLPAPSDFKPIKEE1 179
DB      457 YFPQSEGEDELCTFLANECKTGFCHLYKVTAVLKSQGYDMSSEPFSGEDEFKPIKEE1 516
QY      180 AITSGMEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVVSYNNGEVRRLTDGYS 239
DB      517 ALTSGEWEVLARHGSKIWNVEFTKLVEYFQSTKDTPLEHLLYVVSYEAAGEIVRLTPGFS 576
QY      240 HSCCIGQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCTKKEFMATILDSAGPLPDYTP 299
DB      577 HSCSMQNFDMFVSHYSSVSSTPPCVHYKLSGPDDELHAKQPRFWMASMMEAASCPPDYVP 636
QY      300 PEIFSFEFTGTFLYGLMYKPHDLQPGKKYPTVLFYGGPOVAI 343
DB      637 PEIFHFHTRSDVRLYGMITFKPHALQPGKKHPTVLFYGGPOVQ 680
    
```

Search completed: October 15, 2003, 17:28:51
Job time : 67.704 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 12.2186 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-5
Perfect score: 2482
Sequence: 1 TGTANPKYTFKMSKIMIDAE.....HLHLVGLNLSGRALAKVI 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	462	18.6	931	T32919	hypothetical prote
2	429.5	17.3	738	A87516	dipeptidyl peptida
3	355	14.3	741	JCS142	X-Pro dipeptidyl-P
4	300.5	12.1	711	S66261	X-Pro dipeptidyl-P
5	277	11.2	818	A30107	dipeptidyl aminope
6	267.5	10.8	931	A49737	dipeptidyl aminope
7	267	10.8	793	T41703	dipeptidyl aminope
8	248.5	10.0	792	A39914	dipeptidyl-peptida
9	238	9.6	766	CDH26	dipeptidyl-peptida
10	230	9.3	760	S23752	dipeptidyl-peptida
11	226	9.1	779	T25173	hypothetical prote
12	226	9.1	799	T25174	hypothetical prote
13	221	8.9	803	A41793	dipeptidyl aminope
14	216	8.7	803	T68600	dipeptidyl aminope
15	216	8.7	865	T54331	dipeptidyl aminope
16	176	7.1	829	T19514	hypothetical prote
17	173.5	7.0	795	F82858	dipeptidyl-peptida
18	160.5	6.5	759	T38593	fibroblast activat
19	145.5	5.9	743	T37700	probable dipeptidyl
20	139	5.6	642	C71137	hypothetical prote
21	136.5	5.5	709	B82580	alanyl dipeptidyl
22	134	5.4	683	E87495	prolyl oligopeptid
23	131	5.3	631	H75007	probable acylamino
24	113.5	4.6	466	B89877	hypothetical prote
25	111.5	4.5	1018	SCG82A	collagen alpha 2(IV
26	108.5	4.4	1415	S52267	DNA polymerase III
27	105	4.2	692	S61200	probable membrane
28	104.5	4.2	265	B84063	hypothetical prote
29	104	4.2	1438	C89900	DNA polymerase III

30	103.5	4.2	1835	2	S46082	urea carboxylase (
31	103.5	4.2	4452	1	YC6SG2	gramicidin S synth
32	102.5	4.1	592	2	C70108	peptidase homolog
33	102	4.1	732	1	A35655	peptidyl-dipeptida
34	102	4.1	1312	1	A34171	peptidyl-dipeptida
35	101.5	4.1	4450	2	JX0340	gramicidin S synth
36	901	4.1	1313	1	JC2038	peptidyl-dipeptida
37	100	4.0	558	2	A43667	berine-C-palmitoyl
38	99.5	4.0	543	2	S46098	probable acid-CoA
39	99.5	4.0	591	2	H72474	probable acylamino
40	99.5	4.0	1948	2	B69511	N conserved hypoch
41	97	3.9	866	2	T79267	trial protein - Bac
42	96.5	3.9	440	2	T24073	hypothetical prote
43	96.5	3.9	957	2	H63141	hypothetical prote
44	96.5	3.9	2560	1	I40457	peptide synthetase
45	96	3.9	413	2	D72260	aminotransferase,

ALIGNMENTS

RESULT 1	T32919	hypothetical protein K02F2.1 - Caenorhabditis elegans
C:Species:	Caenorhabditis elegans	
C:Date:	29-Oct-1999	#sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession:	T32919	
R:Maggi, L.; Goela, D.		
submitted to the EMBL Data Library, January 1998		
A:Description:	The sequence of C. elegans comid K02F2.	
A:Reference number:	221246	
A:Accession:	T32919	
A:Status:	preliminary; translated from GB/EMBL/DBJ	
A:Molecule type:	DNA	
A:Residues:	1-931 <MAG>	
A:Cross-references:	AF043699; PIDN:AA897564.1; GSPDB:GN00019; CESP:K02F2.1	
A:Experimental source:	strain Bristol N2; clone K02F2	
C:Genetics:		
A:Gene:	CESP:K02F2.1	
A:Map position:	1	
A:introns:	58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2; 5	
Query Match	18.6%; Score 462; DB 2; Length 931;	
Best Local Similarity	25.6%; Pred. No. 1.2e-28;	
Matches	954; Conservative 79; Mismatches 177; Indels 192; Gaps 23;	
QY	2	GTANPKYTFKMSKIMIDAEGRIVDIDKELIQPFELFEG--VEYIARAGWTREGKVM 58
DB	355	GTNAYSTLRM---VILENGKAYDVPKLD-----EVIKHCFFEYITRAGFFSDGTTVM 406
QY	59	SILDRSQRLQIVLI-----SPELFIVEDDWMERORLIESV 96
DB	407	VQVMSRDQAQCSILLIPYDTPFLPREGLSIKEDNLQSTDLNMGVWD--KSHETMEKP 465
QY	97	PDSVT--PLIYEETTDIWINIHDFHVPQSHSEE--IEFPAECKTGPHLYKITSI 152
DB	466	PRGKLRGTVOQHKARNIDYMINTHNAIYPLKTTDEHPMWEFTYCLEKPNG--SCLLLISA 524
QY	153	LKESKYKSSGGLPAPSDFKCPRIKEIAITSCEWEVLGRHGSNIQVDEVRALVVEFGTD 212
DB	525	LDQNY-----CRHTEKLMENFEI--NKSNGIYVDEVRLEYVYANES 568
QY	213	SPLBHLHYVSVYVNGEVTRLTRNGYSHSCCISOHCDFPISKYSNOK----NPHCVSLYK 268
DB	569	HPTENNI--CVSHYRTGHAQLTESGI-----C--FKSBRANGKLLALDDHGFACVM 616
QY	269	LS--SPEDDPCKTKEF--W-----ATLDSAGP--LPD--YTPPELFSFS--TT 309
DB	617	TSVGSFAE---CRYSFRKENEVLPTVYANITVSHPOQPDHAFSPWEIFQSXT 673
QY	310	GFTLYGMLYKPHDLOPKKYPFTVLFIYGGP----- 339
DB	674	GLMHYAMILRSPNDPYKXVVFHYVGGPGLQIVHNDFSWIQYIRFCRLGVVVFIDNR 733

QY 340 ----- 339
Db 734 GSAHRIEPRRIHKKMGIVVEDQVEGLMLAERTGPFMDNRVVVHGMSYGGYALQM 793
QY 340 -----OVALGAPVTLMTFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391
Db 794 IAKHPNIVRAALIGGAVSDWRLYDTAVTERYMGYP-LEEHVYGASSITGLVEXLPDEPNR 852
QY 392 LILLHGFLENVHFAHTSILSLFLVRAGKPYDLOIYPOEHSIRVPESGEGYELHLHL 451
Db 853 LMLVHGLMDENVHFAHLTHLVDECIRKKGKMHVIFPNERHGVNRNDASIYLDARMYFA 912
QY 452 OE 453
Db 913 QQ 914
RESULT 2
A87516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87516
R:Netman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heideberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolot
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-738 <SNO>
A:Cross-references: GB:AE005673; NID:q13423647; PIDN:AAK24125.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2154
Query Match 17.3%; Score 429.5; DB 2; Length 738;
Best local Similarity 24.9%; Pred. No. 3.6e-26;
Matches 123; Conservative 57; Mismatches 146; Indels 167; Gaps 12;
QY 44 YVARAGTPEGKXAMSLIDRSQTRLOIYLSPELFIPEVDVMERORLIESVDSVTP 103
Db 284 YVARVMSADGKTVYVORLSRDQKTLDLAF-----DATGA 320
QY 104 --IIYETTDIWINIHDIHVPQSHBEIEFIASECTGPHLYKITSILKSKYKS 161
Db 321 GKTIILDTDPHFLEVSNDPRLTDG-----TFLKGE-KDGNQHLRYA-----A 364
QY 162 SGGLPAPSPDFKCPKEIEIAITSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYV 221
Db 365 DGLIL- ----QITKGDMPVIGLEG-----VDEARKVAIFASIDTPIERLYE 408
QY 222 VSVVNGEVTRLTDRCYSHSCCISQHCDFISKYSNOKNHCVSLYLSLP-----EDP 276
Db 409 VSYAKGKPKRALTSAGGWMVAKVADNGAFAGTSDKTPSOALYSADSKRVRMIENK 468
QY 277 TCKTEFWATILDSAGPLPYTPPELIFSFEFTGFTLYGMLYKPHDLOPKKYPVLYFIY 336
Db 469 LAEGHRYW---PYANLPO---PEFSLKAADGETLHTEILRIGDPAPKAFIAYSIV 521
QY 337 GGP----- 339
Db 522 GGFHAGRVKGMHSPSEFTYLEAGYVIFKLDNRGSGNRSAKFMRALDRKLGTEVEDQL 581
QY 340 -----OVALGAPVTLMTFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391
Db 582 GAKFLASQPYVDADKLGVKMSYGGFVALMLTAENTPFRAAGAAPTMSLYDTATYE 641
QY 361 RYMGHPDNEGGYLLGVAMQAEKFPSEPNRLLHGFLENVHFAHTSILSLFLVRAG 451
Db 642 RYMGKPDENAGYAYSDINNRIKTL--APGSLILHGMADDNVIFENSTRLMALQKAI 639

QY 421 PYDLOIYPOEHS 433
Db 700 LPEFAMYPGERHS 712
RESULT 3
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C:Accession: JC5142
R:Kashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expresic
A:Reference number: JC5142; MUID:97164011; PMID:9010758
A:Accession: JC5142
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <KAB>
A:Cross-references: DDBJ:DB3263; NID:q1753196; PIDN:BA11872.1; PID:q1753197
C:Comment: This enzyme catalyses the hydrolytic removal of amino-terminal dipeptidyl re
C:Keywords: dipeptidyl-peptidase IV
F:4-18/Domain: transmembrane #status predicted <Trm>
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:717/Active site: His #status predicted
Query Match 14.3%; Score 355; DB 2; Length 741;
Best local Similarity 22.2%; Pred. No. 3.3e-20;
Matches 123; Conservative 77; Mismatches 164; Indels 190; Gaps 17;
QY 2 GTANPVYTFPMSEIMDAEGRITVDIKELIQFELLFBEVEYIARAGTPEGKXAMSL 61
Db 257 GDANVGVKGVSPABOQTOWIDLKEDPT-----YLARNMRDPQHLISQ-R 304
QY 62 LDRSQTRLOIYLSPELFIPEVDVMERORLIESVDSVTPPLIYETTDIWINIHDIH 121
Db 305 QSDQDKKLDLVEVTL-----ASNQR-----VLAHETSPVPLNHSLR 343
QY 122 VFPQSHBEIEFIASECTGPHLYKITSILKSKYKSSGGLPAPSPDFKCPKEIEIAI 161
Db 344 FL-----DQGSILMSSE-RTGFQHLRYI-----DSGKKA-----AL 374
QY 182 TSGEWEVLGRHGSNIQVDEVRLVYFEGTKDSPLEHLLYVSVNGEVTRLTDRCYSHS 241
Db 375 THGNMSV---DELLAVDEKAGLAYFRAGIESARESOIYAVP-LOGGQPORLKAQGMHS 429
QY 242 CCISQHCDFEISKYSNOKNHCVSLYLSPEDDPTCKTEFWATIL--DSAGPLPYT- 298
Db 430 ASFARNASVYVDSWNSNSTPQIEPFRANG-----EKIATLVENDLADPKHPYAR 479
QY 299 -----PELIFSFEFTGFTLYGMLYKPHDLOPKKYPVLYFIYGP----- 339
Db 480 YREAPRPVEGTLTAADGKTPLYSVIKAGFPDARKYVAVVYVGGPASQVTDGSPGR 539
QY 340 -----OVALGAPVTLMTFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391
Db 540 GDHLFNYQYLAQGVVYFSLDNRGTPRGRDFGALYKGGQTVVAAQLRGVAMLKQOPV 599
QY 343 -----IAGAPVTLMTFYDTGTYTERYMGHPDNEGGYLLGVAMQAEKFPSEPNR 391
Db 600 DPARIQVQSGWNGCYMTMLAKASDSYACGVAGAVTDMGLDHSYTERYMDLPRANDA 659
QY 372 GYLLGVAMQAEKFPSEPNRLLHGFLENVHFAHTSILSLFLVRAGKPYDLOIYPOER 451
Db 660 GYREARVLTHTIEELRSP---LILHGMADDNVIFENSTRLMALQKAI 639
QY 432 HSIRVPESGEHYEL 445
Db 717 HGLSGADALHRYRV 730

RESULT 4

X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
C:Species: Flavobacterium meningosepticum
C:Date: 28-Oct-1996 #sequence #revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66261
R:Kabashima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
Arch. Biochem. Biophys. 320, 123-128, 1995
A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from *Flavobacterium meningosepticum*
A:Reference number: S66261; MUID:95314307; PMID:7793970
A:Accession: S66261
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-711 <XAB>
C:Cross-references: EMBL:DA2121; NID:G577283; PIDD:BA07702.1; PID:G577284.
;Keywords: dipeptidylpeptide hydrolase

Query Match	12.1%	Score 300.5	DB 2	Length 711
Best Local Similarity	22.0%	Pred. No. 7.3e-16		
Matches 101	Conservative	60	Mismatches 141	Indels 157
				Gaps 11
QY	105	IYEETDIWINIHDI FHVFPQSHEEIEFIFASCEKGTGRHLHYKITSILKESKYKRSSGG	164	
Db	302	LFETEDNAMIEITDNLTWEF-----LDDNSFLWASE-RDGRHLHYWDAAGKRLKK-----	349	
QY	165	LPASDPKCPKKEELAITSGEMVELGRGNSIQYDEVARLYVFEFGTDSPLHHLHYVS-	223	
Db	350	-----QVSKGDMELIINYYGNPKTYE-----YYIDTTEGSGINRK--VWSK 387		
QY	224	-YVNPGEVTRLTDRGYSHSCCISQHCDFEISKYSNQKNPHCVSLYKLSPE-----DD 275		
Db	388	LNINTGTQLTSLNAGNNSAASFSTPFNYFITNTSTAKPTKYILKDNANGKVELQUNDD 447		
QY	276	PTCKTKEFMATILDSAGLPDYTRPELIFSFSTTGFLLYGLYKPHDLOPCKKYPYTLFI 335		
Db	448	LNLNLKS-----DNFIKKEFTITPAAGDQNNAMWIKPKDFPAKKYVPVEMFO 495		
QY	336	YGGP-----QVA-----	342	
Db	496	YSGGSGQVANSMDGNGIWPFDLAKQGLYLVCDGRGTRGTYYKKVTVKQLGXEIE 555		
QY	343	-----TAGAPVYTLMIYDT 356		
Db	556	DQITAAKMLGNQSYVDSKRIGIFPMWISGGMASLMTKADVPKMGIAVAAYVWTRREYDS 615		
QY	357	GTEREYMGHPDQNEQGYLLGSVAMQAEKFPSEPRNLILHGFLENDYFAHTSILLSFLV 416		
Db	616	IYTERFLPOTPENKDQYDNLSPYYAKL---KKRPLLHQTADNDVNHFFQNSMEFSEALI 672		
QY	417	RAGKPYDLQIYPOERHSIRVSEGEYHMLHLHYLENTL 455		
Db	673	QNKQKQFPMAYPRDKNHSIIGENTRPPOLYEXKNTVILNLT 711		

RESULT 5

dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein YHR28c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
 C:Accession: S46780, A30107
 R:Du, Z.
 submitted to the EMBL Data Library, June 1994
 A:Description: The sequence of *S. cerevisiae* coamid 8082.
 A:Reference number: S46773
 A:Accession: S46780
 A:Molecule type: DNA
 A:Residues: 1-818 <DUZ>
 A:Cross-references: EMBL:U01039, NID:g500689, PIDN:AAB68879.1, PID:g500698, GSPDB:GND00007
 R:Berberts, C.J., Pohlig, G., Rothman, J.H., Stevens, T.H.
 Cell Biol. 108, 1363-1373, 1989

2

A>Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an
A|Reference number: A30107; MUID:89174971; PMID:2647766
A|Accession: A30107
A|Molecule type: DNA
A|Residues: 1-82, 'N', 84-124, 'N', 126-181, 'LRLEET', 189-199, 'N', 201-365, 'DFKRGKRKF', 376-
A|Cross-references: EMBL:X15484
A|Note: the authors translated the codon ACC for residue 572 as Asn
C|Genetics8:
A|Gene: SGD:DAF2, STE13, MIPS:YHR028C
A|Cross-references: SGD:S0001070; MIPS:YHR028C
A|Map position: 8R
C|Superfamily: dipeptidyl-peptidase IV
C|Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacuole
P|30-45/Domain: transmembrane #status predicted <TM>
P|63-79, 110, 139, 392, 421/Binding site: carboxylate (Asn) (covalent) #status predicted
P|63-79, 110, 139, 392, 421/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match	11.2#:	Score 277:	DB 1:	Length 818:
Best Local Similarity	21.0#:	Pred. No. 6.8e-14:		
Matches 121:	Conservative 64:	Mismatches 170:	Indels 220:	Gap 22
Qy	48	AGWTPGKRAMSILDRSQRLLQVLISPELFIPEVDVMERORLIESVPSVPLIYE	107	
Db	251	AMMSPTGDYLAFLMKIDEEVGEPII-----PYVODEKDIYEMKSIK-YPKSGTF-----N	301	
Qy	108	ETDIWI-NIHD--IFWVFPQSHHEEELFI-----ASCKTGFRHLKYIT	150	
Db	302	PHALWVMSMDKGSFPHRISGNKKDSSLTEVTWNGNVLYKTTDRSSDILTVFLID	361	
Qy	151	SILSESKYKR---SSSG-----LPADSKCPIKEI-----	179	
Db	362	TIATISVNVNRHESNGGMWETHTNTLFIPIANETDRPHNGVVDLPIQGNHLLAFENS	421	
Qy	180	-----AITSGEWEVLGRHGSNIQVDEVRLLYFECTDPSLEHLVYVSVYNPGEVTRLT	234	
Db	422	SSHYKTLTEGKMEV-----NGPLADSEMENRLYFISTKSSTERHVVYIDLRSPELLEVT	478	
Qy	235	DRGVS--HSCISIQHOCDFISKYSNQNNPH-----C-----VSLYKLSSP	272	
Db	479	DTSGDGYDVDSFSSGRFRGLTYGPRVYPOKIVDFHSRAKCKDKGNVLKSLYHLEKN	538	
Qy	273	EDDETCKTKEMWATILDSAGLPLPDYPPLEIFSFESTGFTLYG--MLYKPHDLP----	325	
Db	539	E-----VLRKI--LEDYAAPRKSFRRLNCKQBEKFDILVNSIEILPNDE	583	
Qy	326	--GKKYPTVLFIYGP-----	339	
Db	584	TLSDHYEVFFFPAYGGPNSQQVKTFSVGFNEVVASQNALVAVVDGRGTGKODFRSLV	643	
Qy	340	-----QVAIAG-----	345	
Db	644	RDRIGDYEARDOQISAAALYGSLLTFVDPOKISLFGWSYOGVITLTKLEKDGRRHFRYKGSV	703	
Qy	346	APVTLMTFIYDTGYTERYMGHPDQNEOGYLYGSV-----AMQAEKPPSENNRLLLLHGFLDE	401	
Db	704	APVTDMEFYSVYTERYMGHTPOENFQDFGVYESSVANNVTL-----AQANRRLLMHGTCDD	757	
Qy	402	NVRFARHTSILSLFLVRAG-KRYDIQIYPOERHSIR	435	
Db	758	NVRFQNSLKFLDLDLNGVENYDVHVPPDDSHSIR	792	

RESULT 6

dipeptidyl aminopeptidase (EC 3.4.14.-) - Yeast (*Saccharomyces cerevisiae*)
 N;Alternate names: protein OS045; protein YOR219c; protein YOR50-9
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 03-Mar-1994 #sequence revision 03-Mar-1994 #text_change 29-Oct-1999
 C;Accession: A49737; S45451; S60946; S67112; S71721
 R;Santa Anna-A, S.; Herskowitz, I.
 submitted to the Protein Sequence Database, July 1993
 A;Reference number: A49737
 A;Accession: A49737
 A;Molecule type: DMA

A:Residues: 1-931 <GAN>
A:Cross-references: GB:L21944; NID:g347196; PIDN:AAA5119.1; PID:g347197
R:Pama-Arriola, S.S.; Herskowitz, I.
Yeast 10, 801-810, 1994

A>Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase
A:Reference number: S45451; MUID:95066382; PMID:7975897
A:Accession: S45451
A:Molecule type: DNA
A:Residues: 1-931 <ANN>

R:Cross-references: EMBL:L21944; NID:g347196; PIDN:AAA5119.1; PID:g347197
R:Gallison, F., Dujon, B.
Submitted to the EMBL Data Library, October 1995

A>Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome X
A:Reference number: S60938
A:Accession: S60946
A:Molecule type: DNA
A:Residues: 1-931 <GAL>

A:Cross-references: EMBL:X22441; NID:g1050762; PIDN:CA63182.1; PID:g1050771
R:Bayer, J., Fairhead, C., Gallion, L., Gallison, F., Michaux, G., Thierry, A., Dujon, B.
Submitted to the Protein Sequence Database, July 1996

A:Reference number: S67104
A:Accession: S67112
A:Molecule type: DNA
A:Residues: 1-931 <BOY>

A:Cross-references: EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; PID:e252398; PID:g142050D
A:Experimental source: strain S286C
R:Gallison, F., Dujon, B.

Yeast 12, 877-885, 1996

A>Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
A:Reference number: S71713; MUID:96437977; PMID:884505
A:Accession: S71721
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-931 <GAM>

A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CA63182.1; PID:g1050771
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1995
A:Gene: SGD:STE13; YCII1
A:Cross-references: SGD:S0005745; MIPS:YOR219c
A:Map position: 15R

C:Function:
Description: Involved in processing of alpha-factor prepropheromone
Keywords: dipetidylpeptide hydrolase; glycoprotein; transmembrane protein
F.113-/Domain: transmembrane #status predicted <TM>
F.377/Binding site: carbohydrate (Asn) (covalent) #status predicted
F.785, 863, 896/Active site: Ser, Asp, His #status predicted

Query Match 10.8%; Score 267.5; DB 2; Length 931;
Best Local Similarity 21.3%; Pred. No. 4.7e-13;
Matches 114; Conservative 63; Mismatches 164; Indels 195; Gaps 18;

Oy TSPELFPVDDVMERQRLIESP---DSVPILIYEETDI---WI-NIHDFHVFPO 125
 |||::|||::|||::|||::|||::|||::|||::|||
Db 44S ISPDR--RFELTRNRSKILDVKVDIPSSQMILVRNTNSLFNGWIEKTDIISIPK 501

Oy SHEEIR-FIFASECKTGFPHLYKITSLKESEKRSGGLPARSDFCPIKEAIRAS 183
 ::::::::::::|||::|||::|||::|||::|||::|||
Db 502 PELKRDYGCTIDHADSRGSFLFYPTVF-----AKEPQLTK 540

Oy 184 GEWEVLGRH-GSNIQDEVARLVVEFGTKDSPLEHLVVSVY-----NPGE 229
 |::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 541 GNMEVGNGIVGYEVED----TIFFANEIGVNWSQHLYSLTSTJONTFOSSLONP-- 594

Oy VTRLDRGVSHSCCIOSHCOPEFIISKYNOKRP-----HCVSLLYLTSSE 273
 ::::~::~|||||::|::|::|::|::|::|::|::|
Db 595 ---SKKYDPDELSSARVAISKKLGCPDPKVAGPLTVLNVAEIHDSDIIDLTXXDE 650

Oy DDPTCKTEKFWMATILDSAGELPDYTPPEIFSFEFTT---GFTLYGMLYKPHELOEGKKYP 330
     ~~~~~~KFXE-----KIKNYDL P-ITSYKTMTVLDDGVINVI ELIKPALNLNKXKYP 694  
Oy     331 TVLFTYGP----- 339  
     ::|||::|::|::|::|::|::|::|::|::|::|::|::|

```

Db      695 ILNVNIGGPGSGQTFTTKSLAFEDQAVVSGLDVILQLEPRRTGGKGMFSRMAEKLGYM 754
QY      340 -----QVALGAPRTL 350
Db      755 EPRDITEVTKKFIQRNSQCHIDSKIAIWGWSYGFTSLKTVELDNGDTFKYAAVAPVTN 814
QY      351 MIFVDGTYTERYMGHPDNEGYYLYGVAAQAEKFPSEPNRLLLHGFLENVHFAHTSI 410
Db      815 WTLVDSVYTERYHMQPSNHEGYEVRSTIQNFKSPES-LKRLFIYHGFDDNVHQNFR 873
QY      411 LLSFLVRAG-KPYDLQIYPOERHSIRVPESGEHYELHLHYLOENLGSRIALKYI 465
Db      874 LVQDQILNLGLTNYDMHIFPDDSDHSIRYHNAQRIFQKLYYWLRLDPAFAERPNTEVL 929

RESULT 7
T11703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C/Accession: T11703
R/Murphy, L.G; Harris, D.; Bartrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A/Reference number: Z22011
A/Accession: T11703
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-793 <MUR>
A/Cross-references: EMBL:AL031180, PIN:CAA20138.1, GSPDB:GN00066
A/Experimental source: strain 972h-, cosmid C2E11 -chimeric
C/Genetics:
A/Gene: SPAC2E11.08
A/Map position: 1
C/Superfamily: dipeptidyl-peptidase IV

Query Match      10.8%; Score 267; DB 2; Length 793;
Best Local Similarity 25.6%; Pred. No. 4,1e-15;
Matches 100; Conservative 37; Mismatches 127; Indels 126; Gaps 14;

QY      179 IAITSGEWEVLGRIHGSNIQVDEVRLVYFEETKDSPLEHLYVVSYPNGEVRILTDRGY 238
Db      412 IYLTSGAMDY---DGRPHIDGDFGNVYFLATLKDSTERHLYYVS-LDTLEIYGITDNGE 467
QY      239 S---HSCCISGHCDFPISKSNQKPRCVSLYKLSPEPDFTCKTKEFMATILDSAGPLP 295
Db      468 DEGYYSSTSFSPGDFYVLYNHYGPRV---WQELRSTKDXCYCISLENSRLKQO---LS 520
QY      296 DYTPPEI---FSPESTTGFYLYQMLVKRPHLDQGGKYPTVLFYGGP---QVA----- 342
Db      521 SITLPSVEYGYKLTNDIT-FNF--MERRPRNFVDVKKIPLYFLAYGGPGSQVAKLPRVD 577
QY      343 ----- 342
Db      578 FQAVLASHPDEFIWTLIDGSGTGFNGNAFRYSVRHLGSEWSDGQAGKFMADLPFVD 637
QY      343 -----IAGAVTLMIFDYDTGYTERYMGHPRONQGY 373
Db      638 ENHVGINGMSYGYLYLTKLETQDVFSYGMAVAAPVTWRLVDSYTERYMDLPQYNKEG- 696
QY      374 YLGSVAAQAEKFPSEPNRLLLHGFLENVHFAHTSILLSFLVRAG-KPYDLQIYPOERH 432
Db      697 YKNSQIHDIYERF-KQLKRFPFAVHOTGDVDNHFQSHMLMGLNLANCYNDMAVFPDSAH 755
QY      433 SIRVPESGEHYELHLHYLOENLGSRIALK 462
Db      756 SI-----SYHNASISIYHRLSEWIGDALGRI 781

RESULT 8
A339914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
N/Alternate names: GP10; membrane glycoprotein 110K; OX-61
N/Contains: dipeptidyl-peptidase IV, soluble form

```

C:Species: *Rattus norvegicus* (Norway rat)  
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999  
 C:Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781  
 R:Hong, W.; Doyle, D.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987  
 A:Title: cDNA cloning for a bile canalliculus domain-specific membrane glycoprotein of rat  
 A:Reference number: A39914; PMID:88068516; PMID:3479775  
 A:Accession: A39914  
 A:Molecule type: mRNA  
 A:Residues: 1-792 <HON>  
 A:Cross-references: GB:J02997; NID:9204463; PIDN:AAA1272.1; PID:9204464  
 R:Ogata, S.; Maum, Y.; Ikehara, Y.  
 J. Biol. Chem. 264, 3596-3601, 1989  
 A:Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA  
 A:Reference number: A33315; PMID:89123496; PMID:2563382  
 A:Accession: A33315  
 A:Molecule type: mRNA  
 A:Residues: 1-37, 'A', 39-182, 'I', 184-331, 'T', 333-351, 'C', 353-393, 'V', 395-561, 'L', 563-623,  
 A:Cross-references: GB:J04591; NID:9203973; PIDN:AAA1096.1; PID:9203974  
 A:Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332  
 A:Accession: B33315  
 A:Molecule type: protein  
 A:Residues: 1-20; 35-54; 427-443; 505-509; 511-520; 530-538; 593-600; 602-608; 618-627 <HO2>  
 R:McCaughan, G.W.; Wicks, J.E.; Creswick, P.F.; Gorrell, M.D.  
 Hepatology 11, 534-544, 1990  
 A:Title: Identification of the bile canallicular cell surface molecule GP110 as the ectop  
 quence.  
 A:Reference number: A60730; PMID:90228896; PMID:1970322  
 A:Accession: A60730  
 A:Molecule type: protein  
 A:Residues: 28-47, 'XX', 50-53, 55-58 <MC>  
 R:Ogata, S.; Maum, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.  
 Biochemistry 31, 2582-2587, 1992  
 A:Title: Identification of the active site residues in dipeptidyl peptidase IV by affinity  
 A:Reference number: A42203; PMID:92190188; PMID:1347701  
 A:Accession: A42203  
 A:Molecule type: protein  
 A:Residues: 'R', 625-630, 'X', 632-648 <OG2>  
 R:Iwaki-Bgawa, S.; Watanabe, Y.; Fujimoto, Y.  
 Biol. Chem. Hoppe-Seyler 374, 973-975, 1993  
 A:Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl p  
 A:Reference number: S38949; PMID:94128239; PMID:7905271  
 A:Accession: S38949  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 281-302 <IMA>  
 R:Hong, W.; Doyle, D.  
 J. Biol. Chem. 263, 16892-16898, 1988  
 A:Title: Membrane orientation of rat GP110 as studied by in vitro translation.  
 A:Reference number: A31781; PMID:89034185; PMID:3182821  
 A:Accession: A31781  
 A:Molecule type: mRNA  
 A:Residues: 1-40 <HO3>  
 C:Comment: This protein is localized to the bile canalliculus, which is the apical domain  
 C:Superfamily: dipeptidyl-peptidase IV  
 C:Keywords: dipeptidyl-peptidase hydrolase; glycoprotein; homodimer; liver; serine protease  
 F:1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM>  
 F:1-28/Domain: signal sequence #link MATS #status experimental <SIG>  
 F:1-6/Domain: intracellular #status predicted <INT>  
 F:7-8/Domain: transmembrane #status predicted <TMN>  
 F:29-34/Domain: extracellular #status predicted <EXT>  
 F:35-792/Product: propeptide #link MATS #status experimental <PRO>  
 F:83-90, 148, 217, 227, 319, 521, 686/Binding site: carbohydrate (Aam) (covalent) #status pred  
 F:631/Active site: Ser #status experimental  
 F:709, 741/Active site: Asp, His #status predicted  
 Query Match 10.0%; Score 248.5; DB 1; Length 792;  
 Best Local Similarity 21.0%; Pred. No. 1.2e-11;  
 Matches 122; Conservative 66; Mismatches 188; Indels 205; Gaps 23;  
 2 STAPKVTFFKMSSEIMIDAGRIIVIDIELIPEITLREGCYVIRACQTPBCKYASITL 61

DB 258 GAVNPFVAFPI--VNTDSLSTTTTIPMQITAPASVT-TGDHYLCDAWVSEDRIS---- 310  
 QY 62 LDRSGTRQIYVLSBELFIPVEDDMERORLLESVPDSTPLIYEETDININHDIFH 121  
 DB 311 -----LQWLRRLQNY--SYMAICDDVCKNLWNPCPTQEH 343  
 QY 122 V-----FPOSEEIEFIFASECKTFRHLKYKITSILKESKYRSSGGLPA 167  
 DB 344 IETSAITGVNRRPRAEPHFTSDSSFYKIVSKDQYK---ICQKQDK----- 390  
 QY 168 PSDFKCPKIEIAITSGEMVYGRHSGNIOQDEVRLVYFEET-KDSPLEHLLYVVSYN 226  
 DB 391 -----PEODCTFIRKGAWEVI-----STEALTSPLYIISNEYKEMPCGRULYKI----- 435  
 QY 227 PEGVRLTRDRGSHSCCIS-----QHCPF---ISKYSQKQKPHC-----VSLYKLSPPD 274  
 DB 436 -----QLTD--HTNKKCLSCDLNPRCCOYVSLSKEAKKYOLGGRGRLPLTYLHRSTD 488  
 QY 275 DPTCKTEKFMATILDSAGPLPYTPPE---LPSFESTGFTLYGMLYKPHDLPQCKYPT 331  
 DB 489 QKELRVLED-NSALDKM--LQDVQMPKSKLDFVUNETRF-WYQMLPRH-FDKSKRPL 543  
 QY 332 VLFYIGP----- 339  
 DB 544 LIDVYAGFOSQKADAFPMNATYLASTENITVASFDGRSGYQGDKIMHAIKRLGTL 603  
 QY 340 -----QVATAG-----APVTLWTFY 354  
 DB 604 VEDQEARQFLKMGFVDSKQVALWMSYGYVTSMLVSGSGVFKCGIYAVPVSMEY 663  
 QY 355 DPGYERAMG--HPDQNGGYVLGSVMAQKFKPEPRRLILHGFLENVFAHTSILL 412  
 DB 664 DSVYTERYMGLEPTEDNLDHRTNSTVMSRAENF--KQVEYLITHTADNVAFQSAQIS 721  
 QY 413 SFLVAGKPYDLQIYPOERHSIRVPESEGHVELHLHYLOE 453  
 DB 722 KALVDAGVDFQAMWYTDHDHGIASGTAGHTYSHMSHFLOQ 762  
 RESULT 9  
 CDH026  
 dipeptidyl-peptidase IV (EC 3.4.14.5) - human  
 N:Alternate names: cell surface glycoprotein CD26; chymocyte-activating molecule (THAM)  
 C:Species: *Homo sapiens* (man)  
 C:Date: 31-Dec-1993 #sequence revision 23-Aug-1996 #text change 18-Jun-1999  
 C:Accession: S24313; B42408; B61136; S59510; S16154; S59857; S15520  
 R:Maum, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.  
 Biochim. Biophys. Acta 1131, 333-336, 1992  
 A:Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a se  
 A:Reference number: S24313; PMID:92329551; PMID:1352704  
 A:Accession: S24313  
 A:Molecule type: mRNA  
 A:Residues: 1-6, 'I', 8-766 <MIS>  
 R:Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Sepin, C.; Troc, P.; Barbat, A  
 J. Biol. Chem. 267, 4824-4833, 1992  
 A:Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon can  
 IV mRNA levels during cell differentiation.  
 A:Reference number: A42408; PMID:92165847; PMID:1347043  
 A:Accession: B42408  
 A:Molecule type: mRNA  
 A:Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-711, 'G', 713-766 <DAR2>  
 A:Cross-references: GB:M0536; NID:9181569; PIDN:AAA52308.1; PID:9181570  
 A:Experimental source: intestine  
 A:Note: this sequence corresponds with the author's translation  
 A:Accession: A42408  
 A:Molecule type: mRNA  
 A:Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-711, 'G', 713-766 <DAR2>  
 A:Cross-references: GB:M0536; NID:9181569; PIDN:AAA52308.1; PID:9181570  
 A:Note: sequence extracted from NCBI backbone (NCBI:83986, NCBI:83988); this sequence  
 R:Gorrell, M.D.; Ferrero, A.; Chambaud, L.; Rigal, A.; Bonicel, J.; Maroux, S.  
 Gastroenterology 101, 618-625, 1991  
 A:Title: Expression of sucrase-isomaltase and dipeptidyl-peptidase IV in human small int

A:Reference number: A61136; MUID:91317403; PMID:1677636  
A:Accession: B61136  
A:Molecule type: protein  
A:Residues: 1-15, 'X', 17-22 <GOR>  
R:Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.  
Biochem. J. 311, 835-843, 1995  
A:Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a  
A:Reference number: S59510; MUID:96067599; PMID:7487939  
A:Accession: S59510  
A:Molecule type: DNA  
A:Residues: 1-31 <BOB>  
A:Cross-references: GB:S79876; NID:g1195574; PIDN:AA835614.1; PID:g1195575  
R:Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, H.  
J. Immunol. 149, 481-486, 1992  
A:Title: Cloning and functional expression of the T cell activation antigen CD26.  
A:Reference number: I56154; MUID:92325476; PMID:1352530  
A:Accession: I56154  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-436, 'S', 438-766 <TN>  
A:Cross-references: GB:M74777; NID:g180082; PIDN:AA51943.1; PID:g180083  
R:Abbot, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.  
Immunogenetics 40, 331-338, 1994  
A:Title: Genomic organization, exact localization, and tissue expression of the human CD  
A:Reference number: S59857; MUID:95012454; PMID:7927537  
A:Accession: S59857  
A:Molecule type: DNA  
A:Residues: 1-436, 'S', 438-766 <ABB>  
A:Cross-references: EMBL:U13734  
C:Genetics:  
A:Gene: GDB:DP4  
A:Cross-references: GDB:125239; OMIM:102720  
A:Map position: 2q24.3-2q24.3  
A:Intons: 2/3; 32/1; 65/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3  
F:7-28/Domain: transmembrane #status predicted <TN>  
F:29-766/Domain: extracellular #status predicted <EXT>  
F:65,92,150,219,229,281,321,550,665/binding site: carboxydrate (Asn) (covalent) #status  
F:630,708,740/Active site: Ser, Asp, His #status predicted  
Query Match 9.6%; Score 238; DB 1; Length 766;  
Best Local Similarity 21.1%; Pred. No. 8.3e-11;  
Matches 120; Conservative 71; Mismatches 193; Indels 186; Gaps 24;  
DB 260 GAVNPVVKFV--VNTDSSSVNATSIOITAPASMLI-GDHYLCVYTA----- 306  
QY 62 LRSQTRLOVLSPFLFVDDVMERORLIESVPSVTPPLIYEETDIIWINDIFR 121  
DB 307 ---TORISLQWL-----RRIGNY--SVMDICDYDESSGRW-NC----- 339  
QY 122 VFPQSHBEERLFFASECKTGFRH-----LYKITSILKSKYKSSGGLPAPSPKCP 174  
DB 340 LVARQHIEMSTGCVGRFRSEPHFTLDGNSFYKIIS--NEEGYRII-----CYQID 390  
QY 175 IKEEIAITSGEVEVLGRHSGNIQVDEVRRLVYEGT-KDSPLRHLVYVSVNGEYTRL 233  
DB 391 KKDCTITKGTWVIG-----EALNSDYLYISNEKMGPGGRNLKYLQIDYTVKLC 445  
QY 234 TDRGSHSCIS-QHCDPFIISKYSNQKNPH--C---VSLYKLSPEDDPTCKTEFMA 285  
DB 446 -----SCENLPERCOYVSFSKEAKYQLRCGPGRLPYTLHSSVNDKGRVLE-D 497  
QY 286 TLDSAGRLPDYRPE---IFSFSTTGFTLYGMLKKPHDLOGKRYPTVLYGCP--- 339  
DB 498 SALDKM--LQNVOMPSKGLDFTILNETKF-WYQMILPPH-FDKSKYRPLLDVYAGPCSQ 553  
QY 340 ----- 339  
DB 554 KADTVRLMAYLASTENIVASFDGRSGSGYGDKIMAHINRLGTFEVEDQIEAARQ 613

QY 340 -----QVAIGAPVTLWIFPDYGTERTMG- 364  
DB 614 SKMGFVNDRIAIWGSYGYVTSMWLSGSGVFKGIAPVSRREYDSVYTERMGL 673  
QY 365 -HPDQEGYLYGSVAMQAEKPEEPERNLLHLGFLDENVHPAHTSILSLVLRACKPYD 423  
DB 674 PFPEDLDHRYNSTVMSRAENF--KQEVYLLHGTADNVHFGQSAQISKALVDGVDFQ 731  
QY 424 LQIYPERHSIRVPESEGEYELHLHYLOE 453  
DB 732 AMWYTDHGHGASSTAHQHIYTHMSHFIRQ 761  
RESULT 10  
523752  
dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse  
N:Alternate names: CD26 alpha subunit; THAM alpha subunit  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999  
R:Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.  
J. Biol. Chem. 267, 2200-2208, 1992  
A:Title: CDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-  
A:Reference number: S23752; MUID:92129288; PMID:1370813  
A:Accession: S23752  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-760 <MAR>  
A:Cross-references: EMBL:X58384  
R:Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;  
J. Immunol. 147, 447-454, 1991  
A:Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase  
A:Reference number: A46465; MUID:9102787; PMID:1712807  
A:Accession: A46465  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 4-20 <WV>  
A:Experimental source: M14.T thymoma cells, Swiss nu/nu  
R:Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.  
Biochemistry 33, 15204-15214, 1994  
A:Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.  
A:Reference number: A56030; MUID:95092780; PMID:799781  
A:Accession: A56030  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 746-760 <BER>  
A:Cross-references: GB:U12620  
C:Genetics:  
A:Gene: CD26  
C:Superfamily: dipeptidyl-peptidase IV  
C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein  
F:213,223,315,514,679/binding site: carboxydrate (Asn) (covalent) #status predicted  
F:624,702,734/Active site: Ser, Asp, His #status predicted  
Query Match 9.3%; Score 230; DB 1; Length 760;  
Best Local Similarity 22.6%; Pred. No. 3.6e-10;  
Matches 98; Conservative 47; Mismatches 120; Indels 168; Gaps 19;  
DB 140 KGFRLHYITSLKSKYKSSGGLPAPSPDK-CPKEEIAITSGEVEVLGRHSGNIQV 198  
QY 372 KQGYKAIHF-----PKDKDCTF-----ITKAMEVI-----SIBA 403  
DB 199 DEVRRLVYEGT-KDSPLRHLVYVSVNGEYTRLTDRGSHSCIS-----QHCDPFI 252  
QY 404 LNSDYLYISNOYKEMPGGRNLKYI-----QLTD--HTNVKCLSCDLNPERCOYTA 452  
DB 253 SKYSNKNPH--C---VSLYKLSPEDDPTCKTEFATILDSAG---PLPDYRPE- 301  
QY 453 VFSKSKAKYQYLGCMGPGRLPYTLHRSYTHKELRVLE-----DNSALDRMLQDVGMPK 506  
DB 302 ---IFSFSTTGFTLYGMLKKPHDLOGKRYPTVLYGCP----- 339



```

Db      507 KLDFFVLNMETRR-WYQMLPRH-FDKSKRTPLLDDVYAGPCGKADASFRLKMATYLAST 564
OY      340 -----OVALAG-- 345
Db      565 ENIVASFDRGSGGYQGDKIMHAINRRLGLEVEDOIEAROFVKMGFDVDSKVALTWGMS 624
OY      346 -----APVTLMI FYDTGTERVMG--HPDONEQGYLLGSVAM 380
Db      625 YGGVYTSMWLGSGSGVFCKGIAVAPVSRREYDVSVYTERVMGLPIPEDNLDHYRNSTVMS 684
OY      381 QAEKRPSEPNRLLHLHGFLDENVHFMTSILLSFLVRACKPXYLOIYPOERHSIRVPEEG 440
Db      685 RAENH--KQVELLHGTADNVNHFQGSQIISKVLVDAGVDFOAMMYTDEDHGIASSTAH 742
OY      441 EHYELLHLHYLOE 453
Db      743 OHIYSHMSHFLQO 755

```

## RESULT 11

hypochlorite protein T23F1.7a - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C:Accession: T25173  
 R:Wilkinson, J.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19990  
 A:Accession: T25173  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-779 <NII>  
 A:Cross-references: EMBL:Z81129; PIDN:CAB03411.1; GSPDB:GN00023; CESP:T23F1.7a  
 A:Experimental source: clone T23F1  
 C:GeneticSB:  
 A:Gene: CESP:T23F1.7a  
 A:Map position: 5  
 A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1  
 A:Superfamily: dipeptidyl-peptidase IV

[illegible]

```
Qy      405 FAHTSILSLFVRAGKPYDLQIYVQRHSIRVPESGEHYELLHLYLQE 4533
          |::|::|   |::|::|::|   |   |   |::|::|
Db      727 FONSATILDELQNRGVDFDLMYENQAHSLSRTS--HVGGKTHFLRQ 7733
```

## RESULT 12

hypochemicalaprotein T23F1.7b - *Caenorhabditis elegans*  
 C.Species: *Caenorhabditis elegans*  
 C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C.Accession: T25174  
 R.Wilkinson, J.  
 submitted to the EMBL Data Library, October 1996  
 A.Reference\_number# 219930  
 A.Accession: T25174  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-799 <WIL>  
 A.Cross-references: EMBL:281129, PIDN:CA803412.1, GSPDB:GN00023, CESP:T23F1.7b  
 A.Experimental\_source: clone T23F1  
 C.Genetic8:  
 A.Gene: CESP:T23F1.7b  
 A.Map\_position: 5  
 A.Introns: 13/3, 52/1, 111/3, 197/3, 366/3, 524/1, 557/1, 648/2, 703/1, 743/1  
 A.Superfamily: dipeptidyl-peptidase IV

Query Match	9.1%;	Score 226;	DB 2;	Length 799;
Best Local Similarity	21.1%;	Pred. No. 8e-10;		
Matches 99;	Conservative 62;	Mismatches 142;	Indels 166;	Gaps 18

```

QY 106 YEETDIWIMIHDI PHVFPQSHHEEIEPIFASCKGFRRLYKITSILAKSKYKSSGGL 165
Db 370 YKASAKMWT-HDDFHSI-TSEDTLPFLP-----HDKDNFAQOASLRLSHGOL 419
QY 166 PABDFKCPKIEEIAITSGEMVELGRRGSNIQVDEVRRLVFEGETDSPLENHLVYVSV 225
Db 420 RTEK-----FLNIGEDVTVSINGINKET-----RTIFFMAAPKPBGRHSLSFSYS-- 463
QY 226 NPGEVTRLDRGVSHECCIS---QHCDFTLSKSNQKPHCVSLYKLSSEP-----274
Db 464 -----LADSRKNSAYCISCSIKNCTWMAQMDDMKTAIYSCKPAAPHTAIVNLTRM 516
QY 275 DPTCKTK-----EFWATILDSAGEPLPYTPPEIFSEFSTGFTLYGMLYKPHDLORPK 327
Db 517 DSKKTEHMANLVDKNTYQNRVEAG-LPVIKETI---KISDFDALIKLSIPKOIYNDD 572
QY 328 KY---PTVLFYGGP-----339
Db 573 KHOAIPLIYHVYGGPNDONTKEATQIGIEEVVASASQAALIRIDRGSGCGRWKYRSAYI 632
QY 340 -----QVNIAG-----345
Db 633 GQLCTVBEVEDQIAIKVTLRLYRHLIDARVAAFGWSYGGFMTLSMNVNEAPEOFFKCAVS 692
QY 346 -ABVTLMIPYDTGYTERYMGHPDQNEQGYLGSVAQAQKFEFSEPRRLHLHGFLEDNVH 404
Db 693 VAVVTMPAYYDAYTERBYMG--DAPLEST--SDVTYKLDNFS--TRLLMHGILDNDVH 746
QY 405 FAHTSILSLFLVACKRPYDLOIYPOERHSIRVPESGEYELHLHAYLOE 453
Db 747 FQNSALIBELQNRGVDFDLMVYPNDAHSLSRTS--HVVGKMTHTLRQ 793

```

## RESULT 13

dppx1dyl1 aminopeptidase-like protein (EE3.4.14.-) short form, DPPX-S - bovine  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
 C/Accession: A41793  
 R/Wada, K., Yokotani, N., Hunter, C., Doi, K., Wenthold, R.J., Shimasaki, S.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992  
 A/Title: Differential expression of two distinct forms of mRNA encoding members of a d  
 R/Reference number: A41793; MUID:92108018; PMID:1729689



A:Accession: A41793  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-803 <WAD>  
 A:Cross-references: GB:M76429; NID:g408719; PIDN:AA041623.1; PID:g408720  
 A:Note: Sequence extracted from NCBI Backbone (NCBI:P:75138)  
 C:Superfamily: dipeptidyl-peptidase IV  
 C:Keywords: dipeptidylpeptide hydrolase, glycoprotein  
 F:257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.9%; Score 221; DB 2; Length 803;  
 Best Local Similarity 20.4%; Pred. No. 2e-09;

Matches 107; Conservative 65; Mismatches 178; Indels 174; Gaps 19;

```

QY 44 YIARAGMTPEGKYA--MSILLDRSQRRLQIVLISPELFIPEDDVMERQRLIESVPSDT 101
DB 320 YIMVMKATSTKVAVM---LSRAQN-----VS-----LITLCDAIT 353
QY 102 PLII--YEETTDIWINIHDFHVFPOSHHEIFIFASECKTGFRHLKYITSILK--ESK 157
DB 354 GVCCKHGESEAWL-----HRQNEPVS---KDG-RKFFFRALPQGGQK 397
QY 158 YKSSGGLPAPSPFKCPKEIKELAITSGEVLGRHGSNIQVDEVRLVYEGTKDSPLEH 217
DB 398 FHHITVSSQPNSSNDNIQ---SITGMDVT---KILSYDEKRSQIFLSTEDLPRRR 450
QY 218 HLIVSVYVNGEVTR-----LTDGYSHCSCIQHCDPEISKYSNQKPHCVSLYKLS 271
DB 451 QIYASASTV--GSFNQCLSCDLVNDCTYFSASFSPGADFFLKCCEGQVP--TVSVNTTD 507
QY 272 PEDDPCKTKE-FWATILDSAGLPDYTPPEIFSESTTGFTLYGMLYKPHDLQPKKYP 330
DB 508 KKMFDLETNEHVOKAISRQMPKVEYRKLE---TDDYMLPIQLKPAFTTDTTHAHP 561
QY 331 TVLFYIGCP-----QVALAG-----APVT 349
DB 562 LLLVVGCTGSSQVAEKFVATWETWVSSHGAVVVKCDGRGSGFOGTRLLHEVRRLGL 621
QY 340 -----QVALAG-----APVT 349
DB 622 EEKQMEAVRMLKEPIIDIKTRVAVFGKDYGYLSTYLLPAKGQGAQPVSSCGSALSPIT 681
QY 350 LMIFFYDTGTERVMGHPDQNEQYVIGSVAMQAEKFPSEPNLLHLGHFDENVHFAHTS 409
DB 682 DKRLVASAFSERLGLHGLDNRYEMAKVAHRSAL--EQOQFLVHATADKIHQHTA 739
QY 410 ILLFLVRAGKPYDQIYPOERHSIRVPSGEHYELHLHYLQ 453
DB 740 ELITOLIKGKANYSLQIYDESHYFSSSALQCHLHRSILGFVE 783

```

## RESULT 14

I68600  
 dipeptidyl aminopeptidase like protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: I68600  
 R:Tokocant, N.; Doi, K.; Wenthold, R.J.; Wada, K.  
 Hum. Mol. Genet. 2, 1037-1039, 1993  
 A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat  
 A:Reference number: I54331; MUID:93372805; PMID:8103397  
 A:Accession: I68600  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-803 <RES>  
 A:Cross-references: GB:M96860; NID:g306707; PIDN:AAA35761.1; PID:g306708  
 C:Superfamily: dipeptidyl-peptidase IV

Query Match 8.7%; Score 216; DB 2; Length 803;  
 Best Local Similarity 19.1%; Pred. No. 5.1e-09;

Matches 111; Conservative 79; Mismatches 179; Indels 212; Gaps 23;

QY 2 GTANPKVTPKMSIMIDAGRIIDVIDKELIQFEILFEGVEYIARAGMTPEGKYA--WS 59

```

DB 286 GSENPISILH-----VIGLNGPTHDL---EMPPDDPRMREY-YITWVKATSTKVAVTW- 336
QY 60 ILLDRSQRRLQIVLISPELFIPEDDVMERQRLIESVPSVPLII--YEETTDIWINIH 117
DB 337 -LNRQN-----VS-----LITLCDAITGVCCKHGESEAWL-- 369
QY 118 DIFHVPQSHHEI-----EFIFASECKTGR-HLYKITSILKESKYRSSGGLPAPSD 170
DB 370 -----RQNEPVSQDKRKFFIRALPQGGKGYHTT--VSSQPNSSNDNIQ----- 416
QY 171 FKCPKEIKELAITSGEVLGRHGSNIQVDEVRLVYEGTKDSPLEHLYVSVYVNGEV 230
DB 417 -----SITGMDVT---KILAYDEKGNKIYFLSTEDLPRRRLQYSAN----- 456
QY 231 TLLTDGYSH--SCCISQHCDFPISKYSNQKPHCVSLYKLSR-----EDDPCKTK 281
DB 457 ---TEGNFRQCLSCDLVNCCTYFSASFHSD--FFLLKCEGPGVPVYVNTTDDKK 510
QY 282 EF-----WATILDSAGLPDYTPPEIFSESTTGFTLYGMLYKPHDLQPKKYPVL 333
DB 511 MFDLETNEHVKKAINDRQMPKVEYRDIEL-----DQVNLPMQILKPAFTTDTTHPL 564
QY 334 FYIGCP----- 339
DB 565 VVDGTFGSSQVAEKFVSWETWVSSHGAVVVKCDGRGSGFOGTRLLHEVRRLGLEEK 624
QY 340 -----QVALAG-----APVTLM 352
DB 625 DQMEAVRMLKQIYDRTKTRVAVFGKDYGYLSTYLLPAKENGQOTFGCSALSPITDK 684
QY 353 FYDTGTERVMGHPDQNEQYVIGSVAMQAEKFPSEPNLLHLGHFDENVHFAHTSIL 412
DB 685 LVASAFSERVGLHGLDNRYEMTVAHRSAL--EQOFLIHPTADKIHQHTAELI 742
QY 413 SFLVRAGKPYDQIYPOERHSIRVPSGEHYELHLHYLQ 453
DB 743 TOLIRGKANYSLQIYDESHYFSSSALQCHLHRSILNFVE 783

```

## RESULT 15

I54331  
 dipeptidyl aminopeptidase like protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jul-2000  
 C:Accession: I54331  
 R:Tokocant, N.; Doi, K.; Wenthold, R.J.; Wada, K.  
 Hum. Mol. Genet. 2, 1037-1039, 1993  
 A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat  
 A:Reference number: I54331; MUID:93372805; PMID:8103397  
 A:Accession: I54331  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-865 <RES>  
 A:Cross-references: GB:M96859; NID:g306705; PIDN:AAA35760.1; PID:g306706  
 C:Superfamily: dipeptidyl-peptidase IV

Query Match 8.7%; Score 216; DB 2; Length 865;  
 Best Local Similarity 19.1%; Pred. No. 5.7e-09;

Matches 111; Conservative 79; Mismatches 179; Indels 212; Gaps 23;

```

QY 2 GTANPKVTPKMSIMIDAGRIIDVIDKELIQFEILFEGVEYIARAGMTPEGKYA--WS 59
DB 348 GSENPISILH-----VIGLNGPTHDL---EMPPDDPRMREY-YITWVKATSTKVAVTW- 398
QY 60 ILLDRSQRRLQIVLISPELFIPEDDVMERQRLIESVPSVPLII--YEETTDIWINIH 117
DB 399 -LNRQN-----VS-----LITLCDAITGVCCKHGESEAWL-- 431
QY 118 DIFHVPQSHHEI-----EFIFASECKTGR-HLYKITSILKESKYRSSGGLPAPSD 170
DB 432 -----RQNEPVSQDKRKFFIRALPQGGKGYHTT--VSSQPNSSNDNIQ----- 478

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 (Search time 7.7729 Seconds  
(without alignments)  
2964.152 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482

Sequence: 1 TGTANPKVTFKMSSEIMDAE.....HLHTYQENLGSRIALAKVI 465

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	11.2	818	1 DAP2_YEAST	P18962 saccharomyc
2	267.5	10.8	931	1 STJ3_YEAST	P33894 saccharomyc
3	267	10.8	793	1 YEAB_SCHPO	O14073 schizosacch
4	262	10.6	765	1 DPP4_BOVIN	P81425 bos taurus
5	257	10.4	765	1 DPP4_FELCA	O92317 felis silve
6	249.5	10.1	767	1 DPP4_RAT	P14740 ratius norv
7	229	9.4	766	1 DPP4_HUMAN	P27487 homo sapien
8	224	9.2	760	1 DPP4_MOUSE	P28843 mus musculu
9	221	8.9	863	1 DPP6_BOVIN	P42659 bos taurus
10	218	8.8	804	1 DPP6_MOUSE	O92318 mus musculu
11	216.5	8.7	761	1 SEPR_MOUSE	P97321 mus musculu
12	216	8.7	865	1 DPP6_HUMAN	P42658 homo sapien
13	214	8.6	760	1 SEPR_HUMAN	O12884 homo sapien
14	214	8.6	859	1 DPP6_RAT	P46101 ratius norv
15	145.5	5.9	853	1 YDZF_SCHPO	O92319 schizosacch
16	111.5	4.5	1019	1 CAZ6_HUMAN	P12110 homo sapien
17	104	4.2	1436	1 DPP3_STRAM	O53665 staphylococ
18	104	4.2	1438	1 DPP3_STRAM	O99048 staphylococ
19	103.5	4.2	1335	1 DURI_YEAST	P32328 saccharomyc
20	103.5	4.2	4451	1 GRGB_BACBR	P14688 b gramcidl
21	102	4.1	732	1 ACER_MOUSE	P22667 mus musculu
22	102	4.1	1312	1 ACE_MOUSE	P09470 mus musculu
23	101	4.1	1313	1 ACE_RAT	P47820 ratius norv
24	100	4.0	558	1 LCBI_YEAST	P25045 saccharomyc
25	99.5	4.0	543	1 FAT2_YEAST	P38137 saccharomyc
26	97.5	3.9	984	1 DPOL_NPVAC	P18131 autocographa
27	96.5	3.9	2560	1 PPS2_BACSU	P39846 bacillus su
28	96	3.9	413	1 CSD_THEMA	O94191 thermocoga
29	96	3.9	581	1 PRLR_SHEEP	O46661 ovie aries
30	95.5	3.8	1029	1 CAZ6_MOUSE	O02788 mus musculu
31	95	3.8	692	1 Y650_METUA	O57852 methanococc
32	94.5	3.8	570	1 NCAP_TACV	P18140 tacaribe vi
33	94	3.8	550	1 INVI_HANAN	P40912 hansenula a

34	94	3.8	1014	1 BGAL_BACHD	O9K9c6 bacillus ha
35	93.5	3.8	716	1 YDQA_SCHPO	O13730 schizosacch
36	93.5	3.8	1203	1 SDCL_CAEEL	P24349 caenorhabdi
37	93	3.7	732	1 ACPI_HUMAN	P13798 homo sapien
38	93	3.7	998	1 RRPD_FHV	O66929 flook house
39	93	3.7	1233	1 VLI_REOVD	P15024 reovirus (c
40	92.5	3.7	506	1 GAT4_TREPA	O83983 treponema p
41	91.5	3.7	761	1 YKCI_CAEEL	P41993 caenorhabdi
42	91.5	3.7	1461	1 NEOI_HUMAN	O92859 homo sapien
43	91	3.7	986	1 DPOL_NPVEM	P41712 bombyx mori
44	90.5	3.6	375	1 ADH_GADCA	P26325 gadus calla
45	90.5	3.6	2212	1 RRP1_EBOZM	O05318 ebola virus

## ALIGNMENTS

RESULT 1  
DAP2\_YEAST STANDARD; PRT; 818 AA.  
ID DAP2\_YEAST  
AC P18962;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE dipeptidyl aminopeptidase B (EC 3.4.14.-) (DAP B) (YSCV).  
GN DAP2 OR YHR028C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89174971; PubMed=2647766;  
RA Roberts C.J., Pohlig G., Roehman J.H., Stevens T.H.;  
RT "Structure, biosynthesis, and localization of dipeptidyl  
RT aminopeptidase B, an integral membrane glycoprotein of the yeast  
RT vacuole.";  
RL J. Cell Biol. 108:1363-1373(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
Kucada T., Hillier L., Jier M., Johnston L., Langston Y.,  
Lacaille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,  
Nhan M., Rifkin L., Riles L., St Peter H., Trevas E., Vaughan K.,  
Vaughan M., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII.";  
RL Science 265:2077-2082(1994).  
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE  
CC VACUOLES.  
CC -1- SIMILARITY: Belongs to peptidase family S9B.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: X15484; CAA33512.1; -;  
CC EMBL: U10399; AAB68879.1; -;  
CC PIR: S46780; A30107.  
CC MEROPS: S09.006;  
CC SGD: S0001070; DAP2.  
CC GO: GO:000329; C:vacuolar membrane (sensu Fungi); IDA.  
CC InterPro: IPR002469; DDP1V N term.  
CC InterPro: IPR001375; Peptidase S9.  
CC InterPro: IPR002471; Prol\_endopep\_ser.

DR InterPro: IPR000379; Ser-estrin site.  
 DR Pfam: PR00930; DppIV N-term; 1.  
 DR Pfam: PR00326; Peptidase S9; 1.  
 DR PROSITE: PS00708; PRO-ENDOPEP\_SER; 1.  
 DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;  
 KW Transmembrane; Glycoprotein; Signal-anchor;  
 FT DOMAIN 1 29  
 FT TRANSMEM 30 45  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT LIGANDAL (POTENTIAL).  
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 679 679  
 FT ACT\_SITE 756 756  
 FT ACT\_SITE 789 789  
 FT CARBOHYD 63 63  
 FT CARBOHYD 79 79  
 FT CARBOHYD 110 110  
 FT CARBOHYD 139 139  
 FT CARBOHYD 372 372  
 FT CARBOHYD 392 392  
 FT CARBOHYD 421 421  
 FT CARBOHYD 738 738  
 FT CONFLICT 83 83  
 FT CONFLICT 125 125  
 FT CONFLICT 182 188  
 FT CONFLICT 200 200  
 FT CONFLICT 366 375  
 FT CONFLICT 808 818  
 FT SEQUENCE 818 AA; 93404 MW; 318F45045375BD3 CRC64; .  
 Query Match 11.2%; Score 277; DB 1; Length 818;  
 Best Local Similarity 21.0%; Pred. No. 2.2e-14;  
 Matches 121; Conservative 64; Mismatches 170; Indels 220; Gaps 22;

RESULT 2  
 ST13\_YEAST  
 ID ST13\_YEAST STANDARD; PRT; 931 AA.  
 AC P33894;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dipeptidyl aminopeptidase A (EC 3.4.14.-) (BBAP A) (YSCIV).  
 GN STE13 OR YC11 OR YOR219C OR YOR50-9.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9506382; PubMed=7975897;  
 RA Anna-Attila S.S., Herskowitz I.;  
 RT "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl  
 aminopeptidase.";  
 RL Yeast 10:801-810(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AB320;  
 RA Flanagan C.A., Thorner J.;  
 RT "STE13.";  
 RL (in) Getting M.-J., Novick P., Stevens T.H., Rocheblatt J. (eds.);  
 RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University  
 Press, Oxford (1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=96437977; PubMed=8840505;  
 RA Galisone F., Dujon B.;  
 RT "Sequence and analysis of a 33 kb fragment from the right arm of  
 chromosome XV of the yeast Saccharomyces cerevisiae.";  
 RL Yeast 12:877-885(1996).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE  
 ALPHA-FACTOR PRECURSOR.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE  
 VACUOLAR.  
 CC -1- SIMILARITY: Belongs to peptidase family S9B.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC EMBL: L21944; AAA5119.1; -  
 CC EMBL: U08230; AAA17897.1; -  
 CC EMBL: X82441; CAA63182.1; -  
 CC EMBL: 275127; CAA99437.1; -  
 CC PIR: A49737; A49737.  
 CC MEROPS: S09.005; -  
 CC SGD: S0005745; STE13.  
 CC GO: GO:0005802; C:Golgi trans-face; IDA.  
 CC GO: GO:0041777; P:aminopeptidase activity; IDA.  
 CC GO: GO:007323; P:peptide pheromone maturation; IDA.  
 DR InterPro: IPR002469; DppIV\_N-term.  
 DR InterPro: IPR001375; Peptidase S9.  
 DR InterPro: IPR002471; Prol endopep ser.  
 DR InterPro: IPR000379; Ser-estrin site.  
 DR Pfam: PR00930; DppIV\_N-term; 1.  
 DR Pfam: PR00326; Peptidase S9; 1.  
 DR PROSITE: PS00708; PRO-ENDOPEP\_SER; FALSE NEG.  
 DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;  
 KW Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.  
 FT DOMAIN 1 119  
 FT TRANSMEM 120 140  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

```

FT DOMAIN 141 931 LUMENAL (POTENTIAL). (BY SIMILARITY).
FT ACT SITE 785 785 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 863 863 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 896 896 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 931 AA; 107200 MW; 81AF70094093C023 CRC64;

Query Match 10.8%; Score 267.5; DB 1; Length 931;
Best Local Similarity 21.3%; Pred. No. 1.5e-13;
Matches 114; Conservative 63; Mismatches 164; Indels 195; Gaps 18;

OY 74 ISPELFIPEVDVEMERORLIESVP-----DSVPLLIYEETDI---WI-NHIDIFHVEFQ 125
DB 445 ISPDF---RFEITRNSKILDVKYDIPSSQMLVTRNTNSLFGMWIEKTDILSIPK 501
OY 126 SHEEIE-FTPASCKTGFRHLKYITSLKSKYKRSGLPADSDFKPIKEIATIS 183
DB 502 PELKMDYGDYDIHADSRGFSLFYPTVF-----AKEPITOLTK 540
OY 184 GEMEVLGRH--GSNIQVDEVRRLVFEETKDSPLFHHLYVSVV-----NPG 229
DB 541 GMEVETGNGIYGEYETD---TTFETANEIGVMSOHLISLTDSTONTQSLQNP-- 594
OY 230 VTRLRGVSHSCCISQCHDFPISKYSNKNP-----HCVSLYKLSPE 273
DB 595 ---SDKDYDFDFELSSARAYISKKGDPRIKVAGPLTRVLANVAEIHDSILQITDE 650
OY 274 DDPCTKTEFMATILDSAGPLPDYTPPELFESESTT---GFLYGLMYKPHDLOPKKYP 330
DB 651 ---KFE-----KINQYDLP-ITSYKTVLDDGVAINIYIEIKPAILNEKKYP 694
OY 331 TVLFYGGP----- 339
DB 695 ILVNIYGGPGSOTFTTKSGLAEQAVSGLDVIVLQIEPRGTGKGMSEFRSAREKLGW 754
OY 340 -----OVAIAGAPYTL 350
DB 755 EPROITEVTKKFIQNSQIHDESKIAIMGWSYGGFTSLKTVELDNGDTPFKYAMAVAPTN 814
OY 351 WIFDVTGTERYMGPPDNEOGYIYGVAMQAEKPPSEBNLLHGLDENVFAHSTI 410
DB 815 WTLVDVYTERYMNOPSENEHEGYFEVSTIONKFSFS--LKRLLFYHGFDDNVHIONFR 873
OY 411 LLSFLVRAG-KPYDLOIYPOERHSIRVPSGEGHYELHLHYOENLGSRIALAKYI 465
DB 874 LVDOINLGLNRYDMHIFPDSHSIRYNAQRIYFOKLYWLRDAFAERFNTVTL 929

RESULT 3
YEAR SCHPO STANDARD; PRT; 793 AA.
ID YEAR SCHPO
AC 014073;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative dipeptidyl aminopeptidase C2E1.08 (EC 3.4.14.-).
GN SPACUNKA.08 OR SPACCELL.08
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=972;
RX MEDLINE=21846401; Pubmed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle B.V., Hunt S., Jagers K.,

```

```

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders R., Squares K., Sharp S.,
RA Skelton J., Stimmone M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesli D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpikovski G.V., Usero D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC - SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC vacuoles (By similarity).
CC - SIMILARITY: Belongs to peptidase family S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, AL031180; CAA20138.1; -.
CC PIR, T41703; T41703.
CC GeneDB, SPombe; SPACUNKA.08; -.
CC InterPro; IPR002469; DPEP1V N-term.
CC InterPro; IPR001375; Peptidase S9.
CC InterPro; IPR002471; Prol endopep ser.
CC InterPro; IPR000379; Ser_gestre_site.
CC Pfam; PF00930; DPEP1V N-term; 1.
CC DR Pfam; PF00326; Peptidase S9; 1.
CC DR Prosite; PS00708; PRO ENDOPEP SER; 1.
CC KW Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
CC Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
CC DOMAIN 1 24
CC TRANSMEM 25 45
CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
FT DOMAIN 46 793
FT ACT SITE 647 647 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 722 722 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 755 755 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 793 AA; 91304 MW; 20B70FP97F231463 CRC64;

Query Match 10.8%; Score 267; DB 1; Length 793;
Best Local Similarity 25.6%; Pred. No. 1.4e-13;
Matches 100; Conservative 37; Mismatches 127; Indels 126; Gaps 14;

OY 179 IATSGEMEVLGRGNSIQVDEVRRLVFEETKDSPLFHHLYVSVVNGEVTLRDGY 238
DB 412 IYLTSGAMDVT---DGPRIHIDGDFGNVYFLATLKDSTERHLVYVS-LDTLEIYGITNGE 467

```



```

OY 122 VEPQSHHEE1-----EFIPASECKTGFRHLKYKITSILKESKYKRSGLPAPS 169
DB 345 I-----EISTGWGVRFRPAEPHTSGNS-----FYILIS--NEBGKHI-----C 384
OY 170 DFKCPKIEIAITSGEWEVLGRHNSNIQVDEVARLVYFEGT-KDSELEHLLVYVSVNG 228
DB 385 HFQTDKRNCTFITKGAMEVIG-----IEALTSIDLYIISNEYKMGMBANLVKIQLNDYIT 439
OY 229 EYTRLT-----DRGSHSCCISQHCDFIFISKYSNQNPHCVSLYKLSSEDDPTCTKTFP 283
DB 440 KATCLSCELNPRCOYVSFSQEARVYQLRCSGRLP-----LYTLHNSNNKELAVLE- 494
OY 284 WATILDSAGPLPD-YTPPEIFSESTGTGFTL-YGMLYKPHDLQPKKYVTLFTYICGP-- 339
DB 495 --NNSDLDVLDVQVQPSKLDPIHLHGKTFWQMLPRH-FDKSKYKYLLELVYAGPS 551
OY 340 ----- 339
DB 552 OKADAFRLNWTATYLASTENIIVASPDGRSGYGDKIMHAINRLGTFEVEDQIEATRO 611
OY 340 -----QVATAGAPVTIMIFYDGTTERVNG 364
DB 612 FSRMGFVDDKRIA1WQMSYGVYVTSNVLGAGGVFKGIAVAPVSKMEYVDSVYTERVNG 671
OY 365 --HPDNEGGYLLGSVAQAEKFPSEPNRLLLHGFLENVHFAHTSILSLFVRAKGPY 422
DB 672 LPTPEDNLDYSRNSYVMSRAENF--KQVEYLHIGRADNVHFOQSAQISKALVDAGVDF 729
OY 423 DLQIYPOERHSIRVPSGHEHYELHLYLOE 453
DB 730 QSMWYTDHEDHG1ASSYTAHQHYTHMSHFLKQ 760

RESULT 5
ID DP4_FELCA STANDARD; PRT; 765 AA.
AC 09N217;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipetidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
  antigen CD26).
DN DP4 OR CD26.
OS Fells alvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=peripheral blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
  Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
  activation antigen CD26 homologue.";
RU Immunogenetics 50:366-368(1999).
CC -!- FUNCTION: Removes N-terminal dipeptides sequentially from
  polypeptides having unsubstituted N-termini provided that the
  penultimate residue is proline (By similarity).
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|
  Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
  Xcc is neither Pro nor hydroxyproline.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
  a soluble form (By similarity).
CC -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
  by proteolytic processing (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB023952; BAA92344.1; -.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPPIV_N term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Pro_endopep_ser.
DR InterPro; IPR000379; Ser_estr_site.
DR Pfam; PF00930; DPPIV_N term; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;
  Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765
FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT (MDPP).
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDPP) (BY SIMILARITY).
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
  (POTENTIAL).
FT DOMAIN 1 6
FT TRANSMEM 7 29
FT FT
FT DOMAIN 1 6
FT ACT SITE 765
FT ACT SITE 767
FT ACT SITE 739
FT CARBOHYD 84
FT CARBOHYD 91
FT CARBOHYD 149
FT CARBOHYD 178
FT CARBOHYD 228
FT CARBOHYD 280
FT CARBOHYD 320
FT CARBOHYD 330
FT CARBOHYD 331
FT CARBOHYD 519
FT CARBOHYD 684
SQ SEQUENCE 765 AA; 88213 MW; 3EFC9E8A22B175D9 CRC64;

Query Match 10.4%; Score 257; DB 1; Length 765;
Best Local Similarity 21.8%; Pred. No. 8,3e-13;
Matches 127; Conservative 63; Mismatches 182; Indels 210; Gaps 25;

OY 2 GTANPKVPFKMSEIMDAEGRITVDIKELIOPFELFFGVEIYAGCTPECKYMSIL 61
DB 259 GAANP--TVKLVVITDNLNPTNATSVETPPAAL-TGDYLLCVTVANERIS---- 311
OY 62 LDRSQRLQIVLISPELFIPEVDWNEROLIESVDSVTPLIYETTDIMINIDIFH 121
DB 312 -----LQWLRLQNY--SYMDTRDYNSTGKMS----- 338
OY 122 VEPQSHHEEIEFIPASECKTG-----RHLKYKITSILKESKYKRSGLP 166
DB 339 ---SAAQEHIEW-----STTGWVGRFRPAEPHTSGRNFYKILIS--NEBGKHI----- 383
OY 167 ASDPKCPKIEIAITSGEWEVLGRHNSNIQVDEVARLVYFEGT-KDSELEHLLVYVSV 225
DB 384 --CRFQIDKDKCTFITKGAMEVIG-----IEALTYDLYIISNEYKMGMBGRKLVYKI--- 433
OY 226 NFGVTRRLDRGYSGCCIS-----QHCDFIFISKYSNQNPH---C-----VSLYKLSPE 273
DB 434 -----QLND--YTVACLSCELPKRCOYVSFSKKAQYQLRCSGGLPLYLTHRS 485
OY 274 DDPKTKTEFWATILDSAGPLPDYTPPE---IFSESTGTGFTLYGMLYKPHDLQPKKYP 330
DB 486 NDEEIRLVLED-NSALDKM--LQEVQPSKLDPIHLNFKTF-WYQWMLPRH-FDTSKYP 540
OY 331 TVLFTYICGP----- 339
DB 541 LTIIDVYAGCSQKADAFRLNWTATYLASTENIIVASPDGRSGYGDKIMHAINRLGT 600
OY 340 -----QVATAGAPVTIMIFYDGTTERVNG 364

```



```

Db 601 VVEDQIEAARQFSXMGQFVDDRIKRIAMGWSYGVYTSVNLGAGSGVFKGIAVAVPSMREY 666
Oy 354 YDTGCTERYMGHP-DQNEQGYLLGSVAM-QAEKPPSEPNRLLLHGFIDENVFAHFTSIL 411
Db 661 YDSYVTEKRYMGLPTPPDNLDDYKNSYVMSRAENF--KQVEYLLIHGTADDDVHFQSOAOI 718
Oy 412 LSFVFRAGKXPVDLQIYPOERHSIRPESGGEAYEHLHLHYLQE 453
Db 719 SKALVDAGVDFQAMWYTDDEHDIGASGPAHQHITHTMSHFIRQ 760

RESULT 6
DPP4 RAT
ID _DPP4 RAT STANDARD; PRT: 767 AA.
AC P14740.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) ('T-cell activation
DE antigen CD26) (gp110 glycoprotein) (Bile canaliculus domain-specific
DE membrane glycoprotein).
DE DPP4 OR CD26
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89123496; PubMed=2563382;
RA Ogata S., Miumi Y., Ikehara Y.;
RT "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT its cDNA and identification of the NH2-terminal signal sequence as
RT the membrane-anchoring domain."
RL J. Biol. Chem. 264:3596-3601(1989).
RN [2]
RP SEQUENCE FROM N.A..
RX MEDLINE=88068516; PubMed=3479775;
RA Hong W., Doyle D.;
RT "cDNA cloning for a bile canaliculus domain-specific membrane
RT glycoprotein of rat hepatocytes."
RL Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
RN [3]
RP SIGNAL-ANCHOR.
RX MEDLINE=90335089; PubMed=1974258;
RA Hong W., Doyle D.;
RT "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT rat dipeptidyl peptidase IV."
RL J. Cell Biol. 111:323-328(1990).
RN [4]
RP SEQUENCE OF 281-302.
RX TISSUE=Kidney;
RA MEDLINE=94128239; PubMed=7905271;
RA Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;
RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT dipeptidyl peptidase IV."
RL Biol. Chem. Hoppe-Seyler 374:973-975(1993).
RN [5]
RP FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM THE
POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
PENULTIMATE RESIDUE IS PROLINE.
RN [6]
RP CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
Xcc is neither Pro nor hydroxyproline.
RN [7]
RP SUBUNIT: Homodimer.
RN [8]
RP SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
A SOLUBLE FORM.
RN [9]
RP -1- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
BY PROTEOLYTIC PROCESSING.
RN [10]
RP -1- SIMILARITY: Belongs to peptidase family S9B.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC

```

[illegible]

Db 478 LPLYLHRSIDQKEHRLVED-NSALDKM--LQDVMPKSLDPIVLNTRF-WYQMLIPP 533  
 Qy 321 HDLQGRKXPTVLFYGGP----- 339  
 Db 534 H-FDKSKRYPLLDIVAGCSQKADAFRLNATYLASTENIIVASFDGRSGYOGDKIM 592  
 Qy 340 -----OVAIAG----- 345  
 Db 593 HAINRGLTLEVEDQIEARQFLKMGFVDSKRVATIMGMSYGGVYTSMTLGGSGGVFKCGI 652  
 Qy 346 --APVTLFIPTDGTTERYMG--HPDQNEQGYLLGSVAMQAEKFPSEPNRLLLHGFIDE 401  
 Db 653 AVAPSRMEYDYDSVYTERYMGLPTEPDNLHDYRNSYVMSRAENF--KQVEYLLIHGTADD 710  
 Qy 402 NVHFPHSTLSLFLVRACKPYDLQIYPOERHSIRPSESEHEHLLHVLQOE 453  
 Db 711 NVHFQOASQISKALVDAGVDFOAMWYTTDEHGLASSTAHQHYHSHMSHFLQO 762  
 RESULT 7  
 DPP4\_HUMAN STANDARD; PRT; 766 AA.  
 ID \_DPP4\_HUMAN STANDARD; PRT; 766 AA.  
 AC P27487;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (TSP103) (Adenosine deaminase complexing protein-2) (ADABP).  
 GN DPP4 OR ADCP2 OR CD26.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=92329551; PubMed=1352704;  
 RA Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.,  
 RT "Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine proteinase on the cell surface.";  
 RL Biochem. Biophys. Acta 1131:333-336(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95012454; PubMed=7927537;  
 RA Abdoet C.A., Baker E., Sutherland G.R., McCaughan G.W.;  
 RT "Genomic organization, exact localization, and tissue expression of the human CD26 (dipeptidyl peptidase IV) gene.";  
 RL Immunogenetics 40:331-338(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Peritoneal blood;  
 RX MEDLINE=92325476; PubMed=1352530;  
 RA Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,  
 RA Dahlberg H.N., Schlossman S.F., Morimoto C.;  
 RT "Cloning and functional expression of the T cell activation antigen CD26.";  
 RL J. Immunol. 149:481-486(1992).  
 RN [4]  
 RP ERRATUM.  
 RX MEDLINE=93171637; PubMed=8094732;  
 RA Tanaka T.;  
 RL J. Immunol. 150:2090-2090(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinci P., Prance C.,  
 RA Raha S.S., Loguallano N.A., Peter G.J., Abramson R.D., Millaly S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunatane P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,  
 RA Fahy J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 1-551 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=92165847; PubMed=1347043;  
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,  
 RA Trocort P., Barbat A.;  
 RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines HT-29 and Caco-2. Cloning of the complete human coding sequence and changes of dipeptidyl peptidase IV mRNA levels during cell differentiation.";  
 RL J. Biol. Chem. 267:4824-4833(1992).  
 RN [7]  
 RP SEQUENCE OF 545-766 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=91024044; PubMed=1977364;  
 RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;  
 RT "Isolation of a cDNA probe for the human intestinal dipeptidyl peptidase IV and assignment of the gene locus DPP4 to chromosome 2.";  
 RL Ann. Hum. Genet. 54:191-197(1990).  
 RN [8]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RX MEDLINE=96067599; PubMed=7487939;  
 RA Boehm S.K., Gum J.R., Jr., Erickson R.H., Hicks J.W., Kim Y.S.;  
 RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a TATA-less GC-rich sequence characteristic of a housekeeping gene promoter.";  
 RL Biochem. J. 311:835-843(1995).  
 RN [9]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Kidney;  
 RX MEDLINE=93210468; PubMed=8096237;  
 RA Morrison M.E., Vijayaradhil S., Engelstein D., Albino A.P.,  
 RA Houghton A.N.;  
 RT "A marker for neoplastic progression of human melanocytes is a cell surface ectopeptidase.";  
 RL J. Exp. Med. 177:1135-1143(1993).  
 CC -I- FUNCTION:Removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. Plays a role in T cell activation.  
 CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-Xcc, from a polypeptide, preferentially when Xbb is Pro, provided Xcc is neither Pro nor hydroxyproline.  
 CC -I- SUBUNIT: Homodimer or heterodimer with seprase (FAP).  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in a soluble form.  
 CC -I- PTM: The soluble form (SDPP) derives from the membrane form (MDPP) by proteolytic processing.  
 CC -I- SIMILARITY: Belongs to peptidase family S9B.  
 CC -I- DATABASE: NAME=PRO; NOTE=CD guide CD26 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd26.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)



RP REVISIONS.  
RA Marguet D.A.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B10.A; TISSUE=Liver;  
RA MEDLINE=95092780; PubMed=7999781;  
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;  
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene";  
RL Biochemistry 33:15204-15214(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins B.S., Wagner L., Shenmen C.M., Schlier G.D.,  
RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullan S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smolins D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Merra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP SEQUENCE OF 1-20.  
RX MEDLINE=91302787; PubMed=1712807;  
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,  
RA Bernard A.-M., Gorvel J.-P., Pierres M.;  
RT "Evidence that thymocyte-activating molecule is mouse CD26  
(dipeptidyl peptidase IV)".  
RL J. Immunol. 147:447-454(1991).  
CC -I- FUNCTION: Removes N-terminal dipeptides sequentially from  
CC polypeptides having unsubstituted N-termini provided that the  
CC penultimate residue is proline.  
CC -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-  
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided  
CC Xcc is neither Pro nor hydroxyproline.  
CC -I- SUBUNIT: Homodimer.  
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in  
CC a soluble form.  
CC -I- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)  
CC by proteolytic processing.  
CC -I- SIMILARITY: Belongs to peptidase family S9B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL, X58384; CAA1274.1; -  
CC EMBL, U12609; AAA82213.1; -  
CC EMBL, U12599; AAA82213.1; JOINED.  
CC EMBL, U12600; AAA82213.1; JOINED.  
CC EMBL, U12601; AAA82213.1; JOINED.  
CC EMBL, U12602; AAA82213.1; JOINED.  
CC EMBL, U12603; AAA82213.1; JOINED.  
CC EMBL, U12604; AAA82213.1; JOINED.  
CC EMBL, U12605; AAA82213.1; JOINED.  
CC EMBL, U12606; AAA82213.1; JOINED.  
CC EMBL, U12607; AAA82213.1; JOINED.

DR EMBL, U12608; AAA82213.1; JOINED.  
DR EMBL, U12609; AAA82213.1; JOINED.  
DR EMBL, U12610; AAA82213.1; JOINED.  
DR EMBL, U12611; AAA82213.1; JOINED.  
DR EMBL, U12612; AAA82213.1; JOINED.  
DR EMBL, U12613; AAA82213.1; JOINED.  
DR EMBL, U12614; AAA82213.1; JOINED.  
DR EMBL, U12615; AAA82213.1; JOINED.  
DR EMBL, U12616; AAA82213.1; JOINED.  
DR EMBL, U12617; AAA82213.1; JOINED.  
DR EMBL, U12618; AAA82213.1; JOINED.  
DR EMBL, U12619; AAA82213.1; JOINED.  
DR EMBL, BC022183; AAA82213.1; -  
DR MEROPS; S09\_003; -  
DR MGD; MGI:94919; Dpp4.  
DR InterPro; IPR002469; DPPV\_N\_term.  
DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR002471; Prol\_endopep\_ser.  
DR InterPro; IPR000379; Ser\_exter\_site.  
DR Pfam; PF00930; DPPV\_N\_term; 1.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
DR PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
DR Hydrolase; AminoPeptidase; Dipeptidase; Serine protease;  
DR Transmembrane; Glycoprotein; Signal-anchor.  
DR CHAIN 1 760  
FT CHAIN 37 760  
FT DOMAIN 1 6  
FT TRANSMEM 7 28  
FT FT  
FT DOMAIN 29 760  
FT ACT\_SITE 295 466  
FT ACT\_SITE 624 624  
FT ACT\_SITE 702 702  
FT ACT\_SITE 734 734  
FT CARBOHYD 83 93  
FT CARBOHYD 90 90  
FT CARBOHYD 113 113  
FT CARBOHYD 213 213  
FT CARBOHYD 223 223  
FT CARBOHYD 315 315  
FT CARBOHYD 328 328  
FT CARBOHYD 514 514  
FT CARBOHYD 679 679  
SQ SEQUENCE 760 AA; 87436 MW; ASF644B46A43DF8 CRC64;  
Query Match 9.2%; Score 229; DB 1; Length 760;  
Best Local Similarity 22.6%; Pred. No. 1.5e-10;  
Matches 98; Conservative 47; Mismatches 120; Indels 168; Gaps 19;  
QY 140 KTGPHLYKITSILKSKYKRSSGGLPAPSDPK-CPIKEELAITSGEWELGRHGSNIQV 198  
DB 372 KQGYKHICHF-----PKDKDCTF-----ITGAWEMI-----SIA 403  
QY 199 DEVRLLVYPEGT-KDSPLEHLYVSYNPGVTRLTDRGYSHCIS-----QHCDFIT 252  
DB 404 LTSDLYYISNQYKMPGGRMLYKI-----QLTD--HTNVKCLCDLNPERCQYVA 452  
QY 253 SKYSNQKPH---C-----VSLYKLSPEDDPTCKTEFATILDSAG---PLPDTPP- 301  
DB 453 VSFKEAKYIQLGCGPGPLPYTLTRSTDKELRVL-----DNSALDRMLQDVMPSK 506  
QY 302 --IFSFEFTGFTLYGLMLYKPHDLOPKKKYPTVLFYGGP----- 339  
DB 507 KLDFIVLNETRF-WYQMLPPI-PDKSKYPLLDLVVAGPCQKADASRLMWATYLAST 564  
QY 340 -----QVAIAG-- 345  
DB 565 ENIIASFQSGSGYQGDKIMAINRLGLTEVEDQIEAARQFYKMGFVDSKRVAIWGS 624  
QY 346 -----APVTLMIFVDTGYTERYWG--HPDNOEGYLYGSYAM 380

Db 625 YGGYVTSMLGSSGVFKCIGAVAPVSRMEYDVSYYTERYMGLPIDEDNLHDYRNSTVNS 684  
 QY 381 QAEKFPSEPRRLLLHGFLENVFAHTSTLLSFLVRACKPYDLQIYPOERHSIRVDESG 440  
 Db 685 RAHF--KQVEYLIIHGTADNVHFOQSAISALVDAGVDFOAMMYTDEDHGIASSTAH 742  
 QY 441 EHYELHLHYLOE 453  
 Db 743 QHYSHSHFLQ 755  
 RESULT 9  
 DPP6\_BOVIN  
 ID DPP6\_BOVIN STANDARD; PRT; 863 AA.  
 AC P42659;  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE (Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)  
 DE (Dipeptidyl aminopeptidase 6) (Dipeptidyl peptidase IV like protein)  
 GN (Dipeptidyl aminopeptidase-related protein) (DPPX).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NCBI\_TaxID=9913;  
 [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=92108018; PubMed=1729689;  
 RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimazaki S.;  
 RT "Differential expression of two distinct forms of mRNA encoding  
 members of a dipeptidyl aminopeptidase family.";  
 Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).  
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN  
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE  
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=DPPX-L;  
 CC IsoId=P42659-1; Sequence=Displayed;  
 CC Name=DPPX-S;  
 CC IsoId=P42659-2; Sequence=VSP\_005364;  
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L  
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN  
 CC BRAIN, KIDNEY, OVARY AND TESTIS.  
 CC -!- SIMILARITY: Belongs to peptidase family S9B.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M76428; AAC1622.1; -  
 CC EMBL: M76429; AAC1623.1; -  
 CC PIR: A41793; A41793.  
 DR DR MEROPS: S09.973; -  
 DR InterPro: IPR002469; DPPV\_N\_term.  
 DR InterPro: IPR001375; Peptidase\_S9.  
 DR InterPro: IPR000379; Ser\_estr\_site.  
 DR Pfam: PF00930; DPPV\_N\_term; 1.  
 DR Pfam: PF00326; Peptidase\_S9; 1.  
 DR Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.  
 KM DOMAIN 1 93  
 FT TRANSMEM 94 114  
 FT DOMAIN 115 863  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT EXTRACELLULAR (POTENTIAL).

FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 1 79 MASLYORFTGKINTSRSPAPASRLGGQSEEDGAPK  
 FT PLGAOPAAPAPREPGCGAGGPRFOYOGACDDED ->  
 FT MTTAKENASGKSVQOQEQ (in isoform DPPX-S).  
 FT /FTID=VSP\_005364.  
 SQ SEQUENCE 863 AA; 96556 MM; 23DBA792B841A39D CRC64;  
 Query Match 8.9%; Score 221; DB 1; Length 863;  
 Best Local Similarity 20.4%; Pred. No. 7.7e-10;  
 Matches 107; Conservative 65; Mismatches 178; Indels 174; Gaps 19;  
 QY 44 YIARAGWTPREGKYA--WSILDRSGTRLQIILISPELFIPEVDYMERQRLIESVDSVT 101  
 Db 380 YITWYKMATSTKYAVNM--LSRAQN-----VS-----LITLIDATY 413  
 QY 102 PLII--YEETDIWINIHDIHFVFPQSHHEIEFIASCKTGFRHLKYITSLK--ESK 157  
 Db 414 GCTKKGDESEAWL-----HRQNEPVFS--KQG-RKFFVRALPQCGQCK 457  
 QY 158 YKRSSGGLPAPDFPCPIKEBIAITSGEWEVLGRHGNIOVEVRLVYFEGTKDSPLEH 217  
 Db 458 FHIIVSSSQPRSSNDNIQ---SITSGMDVT---KILSYEKKSQIYFLSTEDLPRRR 510  
 QY 218 HLIVSYVNPGEVTR-----LIDRGYSHSCCISGHCDFISKYSNQNPHCVSLYKSS 271  
 Db 511 QLYSASTV--GSFNNQCLSCDLVDNCTYFSASFSGAFFFLLCKCGPGVP--TVSYNNTD 567  
 QY 272 PEDDPCTKTKF--FMATILDSAGRLPDYPRPELFSESTGFLYGLMYLKHPLDLOGKKYP 330  
 Db 568 KKKMDLEINENVQALISDRQMKVEYKIE-----TDVNLPIQLKPAFTDTATY 621  
 QY 331 TVLFTYGP-----  
 Db 622 LLLVVDGPFSGSVAKEFAVTWETVWSSHGAVVVCPDGRSGFGQTRLHVEVRRRLGSL 681  
 QY 340 -----QVALAG-----APVT 349  
 Db 682 EKKDMEAVRWLKEPYIDKTRVAVFGDGYLSTYLLPAKGDAQPVFSGSALSPIT 741  
 QY 350 LMIFYDTGTERYMGHPONEGGYLLGSVAMQAEKFPSEPRRLLLHGFLENVFAHTS 409  
 Db 742 DPKYASAFSERYLDLHGLDNRAVEMAKVHRVSL--EGQGLVTHATDEKIHFOHTA 799  
 QY 410 ILLSFLVRACKPYDLQIYPOERHSIRVDESGHYELHLHYLOE 453  
 Db 800 ELITQLIKKANYSLOIYFDESHYSSAALQOHLHRSILGFVE 843  
 RESULT 10  
 DPP6\_MOUSE  
 ID DPP6\_MOUSE STANDARD; PRT; 804 AA.  
 AC Q92218; Q9QW2; Q92219;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE (Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)  
 DE (Dipeptidyl aminopeptidase 6) (Dipeptidyl peptidase IV like protein)  
 GN (Dipeptidyl aminopeptidase-related protein) (DPPX).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Embryo;  
 RX MEDLINE=99030650; PubMed=9811881;  
 Hough R.B., Lengeling A., Bedian V., Lo C., Bucan M.;

```

RT      "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
RT      like protein 6 and causes dysregulation of kit expression.";
RT      Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
CC      -I- FUNCTION: May be involved in the physiological processes of brain
CC      function. Has no dipeptidyl aminopeptidase activity. The lack of
CC      activity may be due to the substitution of an aspartate residue
CC      for the serine residue in the proposed catalytic triad (by
CC      similarity).
CC      CC      -I- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC      CC      -I- SIMILARITY: Belongs to peptidase family S98.
CC      CC      -----
CC      CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      CC      the European Bioinformatics Institute. There are no restrictions on way
CC      CC      use by non-profit institutions as long as its content is in no way
CC      CC      modified and this statement is not removed. Usage by and for commercial
CC      CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC      or send an email to license@isb-sib.ch).
CC      CC      -----
DR      EMBL, AF092507, AAC97366.1, -.
DR      EMBL, AF092506, AAC97365.1, -.
DR      EMBL, AF092505, AAC98381.1, -.
DR      MEROPS, S09.973, -.
DR      MGD, MGI:94921, Ddp6.
DR      InterPro, IPR002469, DDPV N term.
DR      InterPro, IPR001375, Peptidase_S9.
DR      InterPro, IPR000379, Ser_centr_site.
DR      Pfam, PF000930, DDPV N term; 1.
DR      Pfam, PF00326, Peptidase S9; 1.
KW      Transmembrane; Glycoprotein; Signal-anchor.
FT      DOMAIN 1 34
FT      TRANSMEM 35 55
FT      FT      EXTRACELLULAR (POTENTIAL).
FT      FT      SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT      FT      (POTENTIAL).
FT      FT      DOMAIN 56 804
FT      FT      CARBOHYD 112 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      FT      CONFLICT 440 455 MISSING (IN REF. 1; AAC98381).
FT      FT      CONFLICT 638 638 Q --> P (IN REF. 1; AAC97365).
SQ      SEQUENCE 804 AA; 91260 MW; 09CFCE7AD8A7168 CRC64;

Query Match 8.8%; Score 218; DB 1; Length 804;
Best Local Similarity 19.0%; Fred. NO. 1.2e-09;
Matches 110; Conservative 78; Mismatches 184; Indels 206; Gaps 22;

Oy      2 GTANPKVTFKMSKSEIADAEGRLIDVIDKELIOPFELFEGVEYIARAGMTEGKYA--WS 59
Db      287 GSENPSTSLH-----VIGANGPTHDL--EMMRPDDPRKREY-YITVMKATSTVYAATM- 337
Oy      60 ILDRSQTRLDIVILSPFLFVEDDVMERQRLIESVDSVTPRII--YEETDIWIIH 117
Db      338 --LRAQKQ--VS-----LITLDDATVGTGCTKHGESESMALH-- 370
Oy      118 DIFHVPDSHEET-----EFIPASEKTEGR-HLYKITSILKESKYKSSGGLPAPSD 170
Db      371 -----KQNEEPFVSKDGRKFFFRALPQGRGKFYHIT--VSSSQPNSNDNIQ---- 417
Oy      171 FKCPILKEIATITSGMEVILGRHGSNIQVDEYRRLVYFEGTQDPSLEHLHYVSVYNGEV 230
Db      418 -----STSGMDVDT--KILSIDERKANKIYFLSTEDLPFRRLRLVSANTVD--DF 462
Oy      231 TR-----LNRDGYSHSCCISQHCDFPISKYSNOKNPICV-----SLTKLSPEDDP 276
Db      463 NRQGLSCDLVENCYTVSASFHNMDFLKEGEGVPIVYHNTTDKRRMDELEANE--- 519
Oy      277 TCKTKERFATILDSAGLPDITTPPEIRSFSTTGFTLYGMILYKPHDIQPGKKYPTVLFIY 336
Db      520 -----EVQKALINDQMPKIEVRKLEIV-----EDYSLPMQILKATGFTDPAHYLLAVD 568

```

Qy	337	GCP	-----	333
Db	569	GPDSQSVTERFEVWETVLVSHGAVVVKCDGRSGFGQTKLQEVNRRRLGFLKXDM	628	
Qy	340	-----	QVALG-----	APVTLWIPYD 355
Db	629	EAVRTMLKEQYIIDKTRVAVFVGQYGGYSTIYLPAKENGQOTTCGSALSPITDFKLYA	688	
Qy	356	TGYTERYNGHPDQNEQGYGLGVSVMQAEKFPSEPNRLLLLGFLDENVHFAMHTSILLSFL	415	
Db	689	SAFSERYLGLHGLDNRBAVEMTKLARVSAI--EDQQLIHATADEKIHFGHTMELITQL	746	
Qy	416	VRAGKPYDLOITYPOERHSIRVPESGEHVELHLIYLQE	453	
Db	747	IKGKANYSLQIYPPESHYFHSVALKQHLRSIIGFVE	784	
RESULT 11				
SEPR_MOUSE				
ID	SEPR_MOUSE	STANDARD	PRT	761 AA.
AC	P97321			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Seprase (EC 3.4.21.-) (fibroblast activation protein alpha) (Integral			
DE	membrane serine protease).			
GN	FAP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxon=10690;			
RN	(1)			
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).			
RC	STRAIN=BALB/c; TISSUE=Embryo;			
RX	MEDLINE=97284459; PubMed=9139873;			
RT	Niedemeyer J., Scallan M.J., Garin-Chesa P., Dalber C., Fiebig H.H.,			
RT	Old L.J., Rettig W.J., Schnapp A.;			
RT	"Mouse fibroblast activation protein: molecular cloning, alternative			
RT	splicing and expression in the reactive stroma of epithelial			
RT	cancer.";			
RN	Int. J. Cancer 71:383-389(1997).			
RP	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Breast;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stepienko M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Ustin T.B., Toshitsuki S., Cangelosi P., Prange C.,			
RA	Raha S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Huily S.W.,			
RA	Villalón D.K., Wuzny D.M., Sodegren E.J., Lu X., Gibbs R.A.,			
RA	Whaley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Blakesley R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smolius D.E.,			
RA	Schmerch A., Schein J.E., Jones S.U.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences.";			
CC	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
CC	-!- FUNCTION: May have a role in tissue remodeling during development			
CC	and wound healing, and contribute to invasiveness in malignant			
CC	cancers.			
CC	-!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and			
CC	type IV collagen, but not native type I or type IV collagen. Does			
CC	not cleave laminin, fibronectin, fibrin or casein.			
CC	-!- SUBUNIT: Homodimer, or heterodimer with DPP4. The hetero-			
CC	mer is			



```

CC inactive (by similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=P97321-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97321-2; Sequence=VSP_005368;
CC Name=3;
CC IsoId=P97321-3; Sequence=VSP_005369;
CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC embryos from day 7-19 and in new-born mice (P1).
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y10007; CAAT1116.1; -.
CC DR EMBL: BC019190; AAT19190.1; -.
CC DR MEROPS: S09_007; -.
CC DR MGD; MGI:109608; Fap.
CC DR InterPro: IPR002469; DPPIV_N_term.
CC DR InterPro: IPR001375; Peptidase_S9.
CC DR InterPro: IPR002471; Prol_endopep_ser.
CC DR InterPro: IPR000379; Ser_estrs_site.
CC DR Pfam: PR00930; DPPIV_N_tcm; 1.
CC DR Pfam: P00326; Peptidase_S9; 1.
CC DR PROSITE; P500708; PRO_ENDOPEP_SER; 1.
CC DR HydroLase; Protease; Serine protease; Transmembrane; Signal-anchor;
CC Glycoprotein; Alternative splicing.
CC FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT (POTENTIAL)
CC FT DOMAIN 26 761 EXTRACELLULAR (POTENTIAL).
CC FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CARBOHYD 49 49 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 92 92 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 99 99 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 227 227 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 314 314 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 679 679 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT VARSPLIC 31 35 Missing (in isoform 2).
CC FT VARSPLIC 31 63 Missing (in isoform 3).
CC FT VARSPLIC 31 63 Missing (in isoform 3).
CC FT CONFLICT 737 737 S -> L (IN REF. 2).
CC FT SEQUENCE 761 AA; 87944 MW; 9174C3ADA213B25 CRC64;
CC -----
CC Query Match 8.7%; Score 216.5; DB 1; Length 761;
CC Best Local Similarity 18.7%; Pred. No. 1.5e-09; Indels 169; Gaps 10;
CC Matches 81; Conservative 57; Mismatches 127;
CC -----
CC 140 KTGFRLLYKTSILKESKYSKSGGLPAPSPDFKPIKEELTALSGEMVELGRHSGNIQVD 199
CC 371 KQGYKIHNY-----KDYENAIQITSGKMEA----- 397
CC 200 EVARLLVFEKGTQDS-----PLEHLLYVSVNVEGEVTRLTRDGYSHSCICSOH 247
CC 398 ----IYIFRYTQDSLFYSSNEFGYRGRNRYIRISIGNSPPSK-----CVTCHL 443
CC 248 ----CDFPISKYSNOKNPHCVSLY-----KLSPEDDPYCKTKKFWATILDSAGPLDPYR 299
CC 444 RKERCQYTYASPFYKAKYALVLYGGLPISTLHDGRTDQEIQVLENNKELENSLNINQL 503
CC 300 PEI-FSFESTGTGTLVGMLYKPHDLPGKKYPTVLPFIYGGP----- 339

```

```

DB 504 PKEVTKKLKDGGLTFYKMLIPPOFDRSKYFLLIQVYGPCOSQSVKFAVNMITYLAS 563
QY 340 ----- 339
DB 564 KEGIVALTVDGRGTAFQGDKFLHAYVRKLYGVEVEDQLTAVRKFIEMGFIDEERAIWGM 623
QY 340 -----OVAIAGAPYTLMTYDGYTRYWGHF--DNEQGYLYGSVA 379
DB 624 SYGQYVSSIALASGTGLFKGCIYAVPVSSWEYASISYSEFNGLPKDNLEHKNSTVM 683
QY 380 MGAEPKPSRNLLHGHFLDENHFAHNSILSLFLVRGKPYDQLQYPOERHSIRVPS 439
DB 684 AAATFRRAND--YLLHGTADNVHFKNSAQIAKALVNAQVDPMQMYSDNNHGISGNS 741
QY 440 GHHYELHLYLQE 453
DB 742 QHNLVTHMTHFLKQ 755
RESULT 12
ID DPP6_HUMAN STANDARD; PRT; 865 AA.
AC P42658;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN DPP6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
RC TISSUE=Hippocampus;
RX MEDLINE=93372805; PubMed=8103397;
RA Yokotani N., Doi K., Wenthold R.J., Wada K.;
RT "Non-conservation of a catalytic residue in a dipeptidyl
RT aminopeptidase IV-related protein encoded by a gene on human
RT chromosome 7."
RL Hum. Mol. Gene. 2:1037-1039 (1993).
CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC ACTIVITY. MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=DPPX-L;
CC IsoId=P42658-1; Sequence=Displayed;
CC Name=DPPX-S;
CC IsoId=P42658-2; Sequence=VSP_005365;
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC -1- SIMILARITY: Belongs to peptidase family S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M96859; AAA5760.1; -.
CC DR EMBL: M96860; AAA5761.1; -.
CC DR PIR: I54331; I54331.
CC DR MEROPS: S09_973; -.
CC DR Genew; HGNC:3010; DPP6.
CC DR MIM: 126141; -.
CC DR GO; GO:0008239; F:dipeptidyl-peptidase activity; TAS.
CC InterPro; IPR002469; DPPIV_N_term.

```





CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling  
 CC during development, tissue repair, and carcinogenesis.  
 CC -1- PTM: N-glycosylated.  
 CC -1- SIMILARITY: Belongs to peptidase family S9B.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U09278; AAB49652.1; -  
 CC EMBL; U76833; AAC51668.1; -  
 CC EMBL; AF007822; AAF21600.1; -  
 CC MEROPS; S09\_007; -  
 CC Genew; HGNC:3590; PAP.  
 CC MIM; 600403; -  
 CC InterPro; IPR002469; DPPV N term.  
 CC InterPro; IPR001375; Peptidase S9.  
 CC InterPro; IPR002471; Prol endopep ser.  
 CC InterPro; IPR000379; Ser\_estr\_site.  
 CC Pfam; PF00930; DPPV\_N\_term; 1.  
 CC Pfam; PF00326; Peptidase\_S9; 1.  
 CC PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
 CC Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;  
 CC Glycoprotein; Alternative splicing.  
 CC DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC DOMAIN 26 760 EXTRACELLULAR (POTENTIAL).  
 CC ACT SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC ACT\_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CC CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC VARSPLIC 1 521 Missing {in isoform 2}.  
 CC /Ptm=VSP\_005367.  
 CC CONFLICT 207 207 P -> A (IN REF. 2).  
 CC CONFLICT 229 229 K -> T (IN REF. 2).  
 CC CONFLICT 354 354 R -> T (IN REF. 2).  
 CC SEQUENCE 760 AA; 87820 MW; A0D34B4801B07EA CRC64;  
 SQ  
 Query Match 8.64; Score 214; DB 1; Length 760;  
 Best Local Similarity 20.24; Pred. No. 2,4e-09; Indels 164; Gaps 13;  
 Matches 87; Conservative 53; Mismatches 127;  
 QY 140 KTGFRHLKYKTLTKESKYSKSSGGLPAPDFKCPKEEIAITSGEWEVGRHSNQVD 139  
 DB 371 KQGYKIHVI-----KDYENAIQITSGKKEAI-----NI--- 400  
 QY 200 EVARLVYFECTKDSPLEHLLVYVYNPGEVRLTRGYSHS--CISQSH-----CDFIS 253  
 DB 401 -----FRVTQOSLFYSSNEFEYPPGRNIYRISIGSYPPSKVCYVCHLRKRCQYYTA 453  
 QY 254 KYSNQKRPCHVSLY---KLSPEDDPTCKTEFMATILDSAGPLD-----YTPPEIFS 304  
 DB 454 SFSDYAKYVALVYVGGIPISLHGDGTDG---IKILENKELENALNNIQLPKEEIK 509  
 QY 305 FEFTGFTLYGMLYKPHDLQPKKKYPTLVFIYGGP----- 339  
 DB 510 KLEVDETILMYKNILPPQPRSKKYPILLIQVYGGPCQSQSVRSFVNMWISYLASKGSMVI 569  
 QY 340 ----- 339  
 DB 570 ALVDGRGTAFOGDKLLYAVYRKLGVVEVDQITAVRKFIEMGEFIDEKRIAIMGMSYGYV 629  
 QY 340 -----QVALAGAVTLMIFFVDGTGYTERVYGHF--DQNEGGYLYGSVAMQAEKF 385

DB 630 SSLALASGTGLPFCGIAVAVPVSSWEYVASVYTERFENGLPKDNLNHYKNSTWARRAEYF 689  
 QY 386 PSEPNRLLLHGFIDENYFAHTSILSLFVLRGAKYVDLIQYQERHSIRPESG---EH 442  
 DB 690 RNVV--YLLHGRADNMYFQNSAQAKALVNAQVFOAMWYSDQNHGL-----SGLSLNH 743  
 QY 443 YELHLHYLQIE 453  
 DB 744 LYTHMTHFLKQ 754  
 RESULT 14  
 ID DPP6\_RAT STANDARD; PRT; 859 AA.  
 AC P46101;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dipectidyl aminopeptidase-like protein 6 (Dipectidylpeptidase VI)  
 DE (Dipectidylpeptidase 6) (Dipectidyl peptidase IV like protein)  
 DE (Dipectidyl aminopeptidase-related protein) (DPPX).  
 GN DPP6  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).  
 RC TISSUE=Brain;  
 RX MEDLINE=92108018; PubMed=1729689;  
 RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimaaki S.;  
 RT "Differential expression of two distinct forms of mRNA encoding  
 RT members of a dipeptidyl aminopeptidase family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).  
 CC -1- FUNCTION: May be involved in the physiological processes of brain  
 CC function. Has no dipeptidyl aminopeptidase activity. The lack of  
 CC activity may be due to the substitution of an aspartate residue  
 CC for the serine residue in the proposed catalytic triad.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=DPPX-L;  
 CC IsoId=P46101-1; Sequence=Displayed;  
 CC Name=DPPX-S;  
 CC IsoId=P46101-2; Sequence=VSP\_005366;  
 CC -1- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME  
 CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS. IN  
 CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.  
 CC -1- SIMILARITY: Belongs to peptidase family S9B.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M76426; AAC42061.1; -  
 CC EMBL; M76427; AAC42062.1; -  
 CC MEROPS; S09\_973; -  
 CC InterPro; IPR002469; DPPV N term.  
 CC InterPro; IPR001375; Peptidase S9.  
 CC InterPro; IPR000379; Ser\_estr\_site.  
 CC Pfam; PF00930; DPPV\_N\_term; 1.  
 CC Pfam; PF00326; Peptidase\_S9; 1.  
 CC Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.  
 CC DOMAIN 1 89 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 90 110 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC DOMAIN 111 859 EXTRACELLULAR (POTENTIAL).  
 CC CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 0

```

FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 75 MASLQRTGKINTSRSPAPRASHLLGGGGPREDAGSKP
FT VARSPLIC 1 75 LGPOAAVAPRRGGAGRRPRFOYARSDCCDEED -> MTT
FT AKERSASGKSVQOODO (in isoform DPPX-S).
FT /FTid=VSP_005356.
SQ SEQUENCE 859 AA; 97301 MW; CE26856D26ED1268 CRC64;

Query Match 8.6%; Score 214; DB 1; Length 859;
Best Local Similarity 18.9%; Pred. No. 2.8e-09;
Matches 109; Conservative 81; Mismatches 182; Indels 206; Gaps 22;

QY 2 GTANPKYTFKMSIIMDAEGRIDVIDKELIQPEILFEQVEYIARAGWTEGKYA--WS 59
DB 342 GSENPISLSLH---VIGLNGPTHDL---EMMPDDPRRREX-YITVMKATSTKVAATM- 392
QY 60 ILDRSOTRLOIYLISPELFPVEDDWERORLIESVDSVTPPLI--YEETDWINIH 117
DB 393 --LNRQN-----VS-----LTLTCDATGYCTKKGHEDSEAMWL-- 425
QY 118 DIFHVPFOSHEEII-----EFIFASECKTGFR-HLYKITSILKESKYKSGGGLAPASD 170
DB 426 -----KQNEEPFSGDKRKFVFRAPQGRGKFTHT--VSSQPNSSNDNIQ--- 472
QY 171 FKCPIKEIATSGSEWEVLGRHGSNIQDEVRLVFEFGTSPLEHLVYVSVNPGEV 230
DB 473 -----SITSGDMVLT-----ELTYDEKRNKLTFLSTEDLPRRRLYLSANTVDDNR 519
QY 231 TRLTDRGYSRSCISQCHDFISKYSNOKMHCYSLYLSP-----EDPCTKTEF- 283
DB 520 QCL-----SCDLVENCYVSAFSFSNMD---FFLLKCEBPGVPTVTHNTTKRRMFD 569
QY 284 -----WATILDAGPLDVTPEPEIFSESTGFTLYGMKYKPHDLOPKKYPTVLFYI 336
DB 570 LEANEQVKAIYDQMKIEYRKIEV-----EDYSLPMQLKATPTDTAHYLLLVLD 623
QY 337 GGP----- 339
DB 624 GTPGSGQSVSEFEVETWETVLVSHGAVVVKCDGRSGFGQTKLHEVRRLGLEKDM 683
QY 340 -----QVALAG-----APVLMIFYD 355
DB 684 EAVRTMLKEOYIDKTRVAVFGKDYGYLYSTYILPAKGENOGQFTCSALSPIDFKLYA 743
QY 356 TGYERYMGRHDQNEGGYLLGSVAMQAEKPFSEFNRLLHLGFDEVNHFHTSILSFL 415
DB 744 SASERYLGLHGLDNRAYEMTKLAHRSAL--EDQQLIATATDEKIHFOHTALITQL 801
QY 416 VRACKPYDLOIYPOERHSIRVSEGEHYELHLHYLOE 453
DB 802 IKGKANYSLOIYPDESHYFHSVALKQHLXRSIIIGFVVE 839

RESULT 15
YDZF SCHPO STANDARD; PRT; 853 AA.
AC 09P7E9; 013720;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DN Putative dipeptidyl aminopeptidase Cl4C4.15c (EC 3.4.14.-).
GN SPAC14C4.15C OR SPAPJ760.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_taxid=4896;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajendream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Faltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Heitroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K.J., Rutter S., Saunders D., Seeger K., Sharp K.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodard J., Volckaert G., Aert R., Robben J., Glymptres B.,
RA Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Mambuti R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nure P.,
RT "the genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
-1 SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
vacuoles.
-1 SIMILARITY: Belongs to peptidase family S9B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162631; CAB83084.1; -.
DR EMBL; Z98596; CAB11208.1; -.
DR GenDB; Spombe; SPAC14C4.15C; -.
DR InterPro; IPR002469; DPPIV N term.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00930; DPPIV N term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW Serineprotease; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 66 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 67 89 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT LUMENAL (POTENTIAL).
FT ACT_SITE 719 853 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 795 795 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 828 828 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 853 AA; 98341 MW; 38450BA50F8304B6 CRC64;

Query Match 15.9%; Score 145.5; DB 1; Length 853;
Best Local Similarity 19.0%; Pred. No. 0.00091;
Matches 116; Conservative 65; Mismatches 179; Indels 249; Gaps 25;

QY 26 VIDKELIQPEILFEQVEYIARAGWTEGKYAMSILDRSOTRLOIYLISPELFPVEDD 85
DB 277 YVEEIIQSSKAVM-----WSPDGNCLSYLSIDSKVPVH-VLPEQOLDSKVEDQ 325

```

```
OY 86 VMERQRLIESVDPSTPLIYEETDOWINIH-----DIFHPPOQ--HEEIEFIFA 136
Db 326 NRVNPFHYSTPKDIPFV-----KLFVNCFTDGESEIEVDSFPLSTQHRVITDVAMA 379
OY 137 SECKTGF----RHLY-KITSILKESKYKSSGGLPAPSDPKPIKEEIAITSG-----E 185
Db 380 GNEKLMFVEVLNGNTERVTSLPDLSRKTTIENTEVS-----BHPILATSSLHKYLS 432
OY 186 WEVLGRHGSNIQVDVRRLVPEGTK-----DSPLEHHLVYVSVYVNGEVRLLTD-- 235
Db 433 FESLG---NLKERYVRQ--VFLSNKKRIAIYELDNF-----VPIYLPVNISFLSDLY 480
OY 236 -----RGYSHSCCISO-----HCDPFIISKYSNOKNPHCVS---- 265
Db 481 LINNTLYFTAISGSPFSRVRLCTKSILISEINIOIGSLFGIKVSNDOYLLVNYLGP 540
OY 266 -----LY-----KLSPEDDPT-----CK-----TKEFWA 285
Db 541 EIPKOFIYSHEDKVSTSNDSKNNLPDSSSTSLGKVKLELCSNLETNEELIITKEKFA 600
OY 286 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLPQKPKYPTVLFYGGPOVA--- 342
Db 601 -----FPSVFFKVIKVNKITAIQEIIRPNPNRKRKYPVTFHLYGAPQSAVLT 648
OY 343 -----IAG----- 345
Db 649 GKEYMDINELMASVYNFLVIKVIDISDVSGOHLFSDSHELLIKSMIELRSYVDTPYI 708
OY 346 -----APVTLWIFVDTGYTERYMGHPDONEG 372
Db 709 DRRHVGIMGWSFGGYLTCLKILENADFIKTGAVAVAPVTDWRYDYDAYSENLLGAYSKQTTA 768
OY 373 YVLGSVAMQAEKFPSEPNRLLLHGFLDENVHFAHTSILSFLVRAG---KPYDLOIYP 428
Db 769 IYDKTAHYHSENF-RKLGGLVLVHGTSDDNVHIENMTQLTAMVEKGVYNYVPF---IVP 824
OY 429 QERHSIRVP 437
Db 825 NANHEFSDP 833
```

Search completed: October 15, 2003, 17:11:32  
Job time : 11.3773 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 | Search time 29.7397 Seconds

(Without alignments)  
4034.822 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTFKMSIMDAE.....HLHYIGENIGSRNALKVI 465

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2482	100.0	465	Q9HBM3	Q9HBM3 homo sapien
2	2447.5	98.6	831	Q9HBM5	Q9HBM5 homo sapien
3	2441.5	98.4	632	Q96UX1	Q96UX1 homo sapien
4	2422	97.6	882	Q9HBM5	Q9HBM5 homo sapien
5	2422	97.6	882	Q9HBM5	Q9HBM5 homo sapien
6	2326	93.7	892	Q9DAG6	Q9DAG6 mus musculu
7	1494.5	60.2	863	Q9WYD8	Q9WYD8 mus musculu
8	1488.5	60.0	862	Q9HBM7	Q9HBM7 mus musculu
9	1478.5	59.6	862	Q9HBM4	Q9HBM4 mus musculu
10	1377.5	55.5	310	Q9HBM4	Q9HBM4 mus musculu
11	1266.5	51.0	628	Q9HBM4	Q9HBM4 mus musculu
12	1187.5	47.8	508	Q9HBM4	Q9HBM4 mus musculu
13	1154.5	46.5	439	Q9HBM4	Q9HBM4 mus musculu
14	860.5	34.7	360	Q9HBM4	Q9HBM4 mus musculu
15	845	34.0	1042	Q9VC20	Q9VC20 drosophila
16	845	34.0	1102	Q9VC19	Q9VC19 drosophila

17	841.5	33.9	1053	5	Q9HBM7	Q9HBM7 homo sapien
18	787	31.7	469	4	Q9HBM4	Q9HBM4 mus musculu
19	719	29.0	312	4	Q9HBM7	Q9HBM7 homo sapien
20	474.5	19.1	167	4	Q9HBM7	Q9HBM7 homo sapien
21	462	18.6	927	5	Q9HBM7	Q9HBM7 homo sapien
22	462	18.6	931	5	Q9HBM7	Q9HBM7 homo sapien
23	449	18.1	552	10	Q9HBM7	Q9HBM7 homo sapien
24	449	18.1	746	10	Q9HBM7	Q9HBM7 homo sapien
25	429.5	17.3	728	16	Q9HBM7	Q9HBM7 homo sapien
26	415.5	16.7	763	16	Q9HBM7	Q9HBM7 homo sapien
27	355.5	14.3	751	16	Q9HBM7	Q9HBM7 homo sapien
28	355	14.3	741	2	Q9HBM7	Q9HBM7 homo sapien
29	346	13.9	757	16	Q9HBM7	Q9HBM7 homo sapien
30	312	12.6	901	3	Q9HBM7	Q9HBM7 homo sapien
31	309.5	12.5	771	3	Q9HBM7	Q9HBM7 homo sapien
32	308.5	12.4	765	3	Q9HBM7	Q9HBM7 homo sapien
33	305	12.3	723	2	Q9HBM7	Q9HBM7 homo sapien
34	305	12.3	723	2	Q9HBM7	Q9HBM7 homo sapien
35	300.5	12.1	711	2	Q9HBM7	Q9HBM7 homo sapien
36	263.5	10.6	730	2	Q9HBM7	Q9HBM7 homo sapien
37	263	10.6	748	13	Q9HBM7	Q9HBM7 homo sapien
38	247	10.0	711	4	Q9HBM7	Q9HBM7 homo sapien
39	239	9.6	796	4	Q9HBM7	Q9HBM7 homo sapien
40	234	9.4	827	16	Q9HBM7	Q9HBM7 homo sapien
41	227	9.1	761	11	Q9HBM7	Q9HBM7 homo sapien
42	226	9.1	769	5	Q9HBM7	Q9HBM7 homo sapien
43	225.5	9.1	803	16	Q9HBM7	Q9HBM7 homo sapien
44	224.5	9.0	755	13	Q9HBM7	Q9HBM7 homo sapien
45	222.5	9.0	745	5	Q9HBM7	Q9HBM7 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q9HBM3 PRELIMINARY; PRT; 465 AA.

AC Q9HBM3  
DT 01-MAR-2001 (TREMUR1, 16, Created)  
DT 01-MAR-2001 (TREMUR1, 16, Last sequence update)  
DT 01-DEC-2001 (TREMUR1, 19, Last annotation update)  
DE Dipeptidyl peptidase 8 (Fragment).  
GN DPP8.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=20467194; PubMed=11012666;  
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,  
Gorell M.D.,  
RT "Cloning, expression and chromosomal localization of a novel human  
RT dipeptidyl peptidase (DPP) IV homolog, DPP8".  
RL Eur. J. Biochem. 267:6140-6150(2000).  
DR EMBL; AF221636; AAC29768.1; -.  
DR MEROPS; S09.016; -.  
FT NON TER  
SQ SEQUENCE 465 AA; 53197 MW; 229399EC0A4FE29CE CRC64;

Query Match 100.0%; Score 2482; DB 4; Length 465;  
Best Local Similarity 100.0%; Pred. No. 6.1e-207;  
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TGTANPKVTFKMSIMDAEGRRIIDVIDELIQFEILFEGVEYIRAGTTPGKYAWSI 60  
Db 1 TGTANPKVTFKMSIMDAEGRRIIDVIDELIQFEILFEGVEYIRAGTTPGKYAWSI 60  
Qy 61 LDRSOTRIQIVLISPELPIVEDDVMERORLIESVDSVTPLIYEETDWINIHIDF 120  
Db 61 LDRSOTRIQIVLISPELPIVEDDVMERORLIESVDSVTPLIYEETDWINIHIDF 120

```

QY 121 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 180
DB 121 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 180
QY 181 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 240
DB 181 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 240
QY 241 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 300
DB 241 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 300
QY 301 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 360
DB 301 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 360
QY 361 RYMGHPONEGGYLYGVAQAEKPPSEPNRLLLHGLDENVFAHTSILSLVYRAGK 420
DB 361 RYMGHPONEGGYLYGVAQAEKPPSEPNRLLLHGLDENVFAHTSILSLVYRAGK 420
QY 421 PYDLQIYPOERHSIRVPESGHEYLHLHYLOENLGSRIALAKVI 465
DB 421 PYDLQIYPOERHSIRVPESGHEYLHLHYLOENLGSRIALAKVI 465

```

## RESULT 2

```

Q8NEM5 PRELIMINARY; PRT; 831 AA.
ID Q8NEM5
AC Q8NEM5;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Similar to dipeptidylpeptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strauberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC030688; AAI30688.1; -
SQ SEQUENCE 831 AA; 95527 MW; 0B2A13A2FE70CB2 CRC64;

```

Query Match 98.6%; Score 2447.5; DB 4; Length 831;  
 Best Local Similarity 90.5%; Pred. No. 1.4e-203;  
 Matches 465; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEVGYIARAGWTPCKYAWSI 60
DB 318 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEVGYIARAGWTPCKYAWSI 377
QY 61 LLDRSQTRLQIVLISPELFIPVEDVMERORLIESVPDSVTPLIYEETDIDMINHDI 120
DB 378 LLDRSQTRLQIVLISPELFIPVEDVMERORLIESVPDSVTPLIYEETDIDMINHDI 437
QY 121 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 180
DB 121 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 180
QY 438 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 497
DB 438 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 497
QY 181 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 240
DB 181 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 240
QY 498 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 557
DB 498 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 557
QY 241 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 300
DB 241 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 300
QY 558 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 617
DB 558 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 617
QY 301 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 360
DB 301 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 360
QY 618 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 677
DB 618 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 677

```

```

QY 341 -----VAIAGAPVTLMIIFYDTGYTERYMGHPDNEQ 371
DB 678 SLGVVVVVINDRSGCHRGKLFEGAFKYKVAIAGAPVTLMIIFYDTGYTERYMGHPDNEQ 737
QY 372 GYVLGSYAMAQAEKPPSEPNRLLLHGLDENVFAHTSILSLVYRAGKPYDLQIYPOER 431
DB 738 GYVLGSYAMAQAEKPPSEPNRLLLHGLDENVFAHTSILSLVYRAGKPYDLQIYPOER 797
QY 432 HSIRVPESGHEYLHLHYLOENLGSRIALAKVI 465
DB 798 HSIRVPESGHEYLHLHYLOENLGSRIALAKVI 831

```

## RESULT 3

```

Q96JX1 PRELIMINARY; PRT; 632 AA.
ID Q96JX1
AC Q96JX1;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein FLJ14920.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Iogaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Makamatsu A., Nakamura Y., Nagahari K., Masuo Y.,
RA Ninomiya K., Iwanaga T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027826; BAB55395.1; -
DR MEROPS; S09.018; -
KM Hypothetical protein.
SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;

```

Query Match 98.4%; Score 2441.5; DB 4; Length 632;  
 Best Local Similarity 90.3%; Pred. No. 3.1e-203;  
 Matches 484; Conservative 0; Mismatches 1; Indels 49; Gaps 1;

```

QY 1 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEVGYIARAGWTPCKYAWSI 60
DB 119 TGTANPKVTFKMSIMIDAGRIIDVIDKELIQPFILFEVGYIARAGWTPCKYAWSI 178
QY 61 LLDRSQTRLQIVLISPELFIPVEDVMERORLIESVPDSVTPLIYEETDIDMINHDI 120
DB 179 LLDRSQTRLQIVLISPELFIPVEDVMERORLIESVPDSVTPLIYEETDIDMINHDI 238
QY 121 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 180
DB 121 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 180
QY 239 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 298
DB 239 HVEPQSHHEEIEIFASECKTGFRHLKYITSIILKESKYKSSGGLPAPSPDFCKPIKEIA 298
QY 181 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 240
DB 181 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 240
QY 299 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 358
DB 299 ITSGEMVLGRHGSNIQVDEVRRLVVEGCTKDSPLHLHYVSVVNGEVTRLTDRGYSH 358
QY 241 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 300
DB 241 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 300
QY 359 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 418
DB 359 SCCISQHCDFEFISKYSNQKNPHCVSLYKLSPPEDPTCKTKEFWATILDSAGPLPDYTPP 418
QY 301 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 360
DB 301 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 360
QY 419 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 478
DB 419 EIFSFESTTGFTLYGMLYKPHDLOPKKYPVLVFIYGGPOVAIAGAPVTLMIIFYDTGYTE 478
QY 341 -----VAIAGAPVTLMIIFYDTGYTERYMGHPDNEQ 371
DB 479 SLGVVVVVINDRSGCHRGKLFEGAFKYKVAIAGAPVTLMIIFYDTGYTERYMGHPDNEQ 538

```

QY 372 GYLLGSVAMQAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQER 431  
DB 539 GYLLGSVAMQAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQER 558  
QY 432 HSIRVPESGEYELHLHYQENLGSRIALAKVI 465  
DB 599 HSIRVPESGEYELHLHYQENLGSRIALAKVI 632

RESULT 4  
Q9HBM5 PRELIMINARY; PRT; 882 AA.  
ID Q9HBM5  
AC Q9HBM5  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Dipeptidyl peptidase 8.  
GN DPP8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20467194; Pubmed=11012666;  
RA Abbott C.A., Yu D.M.T., Woollett E., Sutherland G.R., McCaughan G.W.,  
RA Gorrell M.D.,  
RT "Cloning, expression and chromosomal localization of a novel human  
RT dipeptidyl peptidase (DPP) IV homology, DPP8".  
RL Eur. J. Biochem. 267:6140-6150 (2000).  
DR EMBL; AF221634; AAG29766.1; -  
DR MEROPS; S09.018; -  
DR InterPro; IPR002469; DPPV\_N term.  
DR InterPro; IPR001375; Peptidase\_S9.  
DR InterPro; IPR000379; Ser\_estrs site.  
DR Pfam; PF00930; DPPV\_N term; 1.  
DR Pfam; PF00326; Peptidase\_S9; 1.  
SQ SEQUENCE 882 AA; 101421 MW; AD801C302DB46528 CRC64;

Query Match 97.6%; Score 2422; DB 4; Length 882;  
Best Local Similarity 82.3%; Pred. No. 2.5e-201;  
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKMSIIMDAEGRIIDYIDKELIQPELIFEGVEYIARAGTPEGKAWSI 60  
DB 318 TGTANPKVTFKMSIIMDAEGRIIDYIDKELIQPELIFEGVEYIARAGTPEGKAWSI 377  
QY 61 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPDSVTPLIYEETTDIMINIDIF 120  
DB 378 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPDSVTPLIYEETTDIMINIDIF 437  
QY 121 HVPQSHHEIEIFIPASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEEIA 180  
DB 438 HVPQSHHEIEIFIPASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEEIA 497  
QY 181 ITSGEVEVLGRHGSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGSVTRLTDGYSH 240  
DB 498 ITSGEVEVLGRHGSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGSVTRLTDGYSH 557  
QY 241 SCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGPLPYTPP 300  
DB 558 SCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGPLPYTPP 617  
QY 301 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQ----- 340  
DB 618 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFGVKYFRNLTLA 677  
QY 341 ----- 340  
DB 678 SLGIVVVVVDNRGSGHGLKFEAGAFKYMGOEIIDQVGLQYLASRYDFILDRVGHG 737  
QY 341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEGYLLGSVAM 380

DB 738 WSYGYLISLALMQRSDIFRVALAGAPVTLMIFYDTGYTERYMGHPDQNEGYLLGSVAM 797  
QY 381 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 440  
DB 798 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 857  
QY 441 EHYELHLHYQENLGSRIALAKVI 465  
DB 858 EHYELHLHYQENLGSRIALAKVI 882

RESULT 5  
Q8IMG7 PRELIMINARY; PRT; 882 AA.  
ID Q8IMG7  
AC Q8IMG7  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Similar to dipeptidyl peptidase 8.  
GN  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC040203; AAH40203.1; -  
SQ SEQUENCE 882 AA; 101391 MW; 86C76AF5BCE707F9 CRC64;

Query Match 97.6%; Score 2422; DB 4; Length 882;  
Best Local Similarity 82.3%; Pred. No. 2.5e-201;  
Matches 465; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

QY 1 TGTANPKVTFKMSIIMDAEGRIIDYIDKELIQPELIFEGVEYIARAGTPEGKAWSI 60  
DB 318 TGTANPKVTFKMSIIMDAEGRIIDYIDKELIQPELIFEGVEYIARAGTPEGKAWSI 377  
QY 61 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPDSVTPLIYEETTDIMINIDIF 120  
DB 378 LLDRSQTRLQIVLISPFLPVEDDVMERQRLIESVPDSVTPLIYEETTDIMINIDIF 437  
QY 121 HVPQSHHEIEIFIPASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEEIA 180  
DB 438 HVPQSHHEIEIFIPASECKTGRHLKYKTSILKESKYRSSGGLPAPSDFKPIKEEIA 497  
QY 181 ITSGEVEVLGRHGSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGSVTRLTDGYSH 240  
DB 498 ITSGEVEVLGRHGSNIQVDEVRRLVYFEGTKOSPLEHNLVYVSYNPGSVTRLTDGYSH 557  
QY 241 SCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGPLPYTPP 300  
DB 558 SCCISOHCDPFIISKYSNQNPHCVSLYKLSPEDDPTCKTKEFMATILDSAGPLPYTPP 617  
QY 301 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQ----- 340  
DB 618 EIFSESTTGFTLYGMLYKPHDLQPKKYPVLYFYGGPQVQVQVNNRFGVKYFRNLTLA 677  
QY 341 ----- 340  
DB 678 SLGIVVVVVDNRGSGHGLKFEAGAFKYMGOEIIDQVGLQYLASRYDFILDRVGHG 737  
QY 341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEGYLLGSVAM 380  
DB 738 WSYGYLISLALMQRSDIFRVALAGAPVTLMIFYDTGYTERYMGHPDQNEGYLLGSVAM 797  
QY 381 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 440  
DB 798 QAEKFPSEPNRLLLHGLDENVHFAHTSILSLFLVRAGKPYDLOIYQERHSIRVPSG 857  
QY 441 EHYELHLHYQENLGSRIALAKVI 465



Db 858 EHYELHLHYQENLGSRIAALKVI 882

# RESULT 6

09D4G6 PRELIMINARY; PRT; 892 AA.

AC 09D4G6; 17, Created)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Adult male testis cDNA, RIKEN full-length enriched library,  
clone:493244F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RN NCBI

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
Hasegaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,  
Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
Okazaki Y., Okido T., Owa C., Saito R., Saito R., Sakai C., Sakai K.,  
Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,  
Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
Muramatsu M., Hayashizaki Y.,  
Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

[12]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
"Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";

RT Nature 420:563-573(2002).

RN [13]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=21085660; PubMed=11217851;

RA RIKEN FANTOM Consortium;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

RN [14]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=99279253; PubMed=10349636;

RA Carninci P., Hayashizaki Y.,  
"High-efficiency full-length cDNA cloning.";

RT Mechn. Enzymol. 303:19-44(1999).

RN [15]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=20499374; PubMed=11042159;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
"Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";

RT Genome Res. 10:1617-1630(2000).

RL [16]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA MEDLINE=20530913; PubMed=11076861;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,

RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-Format  
RT sequencing pipeline with 384 multicapillary sequencer.";

RL Genome Res. 10:1757-1771(2000).

DR EMBL; AK016546; BAB30295.2; -;

SQ SEQUENCE 892 AA; 102284 MW; FDE0BDECA4CA346 CRC64;

Query Match 93.7%; Score 2326; DB 11; Length 892;  
Best Local Similarity 78.4%; Pred No. 5, 4e-193;  
Matches 443; Conservative 10; Mismatches 12; Indels 100; Gaps 1;

1 TGTANPKVTFKQSEIMDAEGR11VDKEL1QPFELFEGVEYIARAGTPEGKXAWSI 60  
328 TGTANPKVTFKQSEIYVDAAG11VDKELVQPFELFEGVEYIARAGTPEGKXAWSI 387  
61 LDRSGTRQIVLISELFIPEVDVMEQRILIESPDSVTPPLIYEETTDIWINIHDF 120  
388 LDRSGTRQIVLISELFIPEVDVMEQRILIESPDSVTPPLIYEETTDIWINIHDF 447  
121 HYPQSHHEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFKPIKEEIA 180  
448 HYPQSHHEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFKPIKEEIT 507  
181 ITSGEMEVIGRHSNIQVDEVRLLVFEKTKDPSLEHMLVYVSYPNGEVRLLTDRGYSH 240  
508 ITSGEMEVIGRHSNIQVDEVRLLVFEKTKDPSLEHMLVYVSYPNGEVRLLTDRGYSH 567  
241 SCISGHCPCPFISKYSNQNPHCVSLYKLSSPEDDPTCTKEFWAIIIDLSAGLPRYTPP 300  
568 SCISGHCPCPFISKYSNQNPHCVSLYKLSSPEDDPTCTKEFWAIIIDLSAGLPRYTPP 627  
301 EIFSFESTGTFLYGLMLYKPHDLPQPKKYPYLVFIYGGPQ----- 340  
628 EIFSFESTGTFLYGLMLYKPHDLPQPKKYPYLVFIYGGPQVQVUNNRFKGVKFRILNTLA 687  
341 ----- 340  
688 SLGVVVVVIDNRGSGCHRLGKFEKGFYKMGQIEIDQVEGLQYLAQYDFIDLRVGING 747  
341 -----VAIAGAPYTLMIFYDTGYTERVWGHGDQEGCYGLSSVM 380  
748 WSYGYLSLMLMQRSDIRFVAIAGAPYTLMIFYDTGYTERVWGHGDQEGCYGLSSVM 807  
381 QAEKFPSEPNRLLLLHGFDENVHFAHTSILSLFLVRACKPYDLOIYPOERHSIRVPESG 440  
808 QAEKFPSEPNRLLLLHGFDENVHFAHTSILSLFLVRACKPYDLOIYPOERHSIRVPESG 867  
441 EHYELHLHYQENLGSRIAALKVI 465  
868 EHYELHLHYQENLGSRIAALKVI 892

Db 858 EHYELHLHYQENLGSRIAALKVI 892

Db 858 EHYELHLHYQENLGSRIAALKVI 892

RESULT 7

08WXD8 PRELIMINARY; PRT; 863 AA.

AC 08WXD8; 20, Created)

DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Dipeptidyl peptidase 9.

GN DPP9.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCBI\_TaxID=9606;

RN NCBI

RP SEQUENCE FROM N.A.

RA Olsen C., Wagtmann N.,  
"Identification and characterization of a novel member of the  
dipeptidyl peptidase IV-related family.";

RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

RL [12]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strauberg R.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF52102; AAL47179.1;  
 DR EMBL; BC037948; AAH37948.1;  
 DR MEROPS; S09.019;  
 DR InterPro; IPR00469; DDPV N term.  
 DR InterPro; IPR001375; peptidase\_S9.  
 DR InterPro; IPR00379; Ser\_estra\_site.  
 DR Pfam; PF00930; DDPV\_N\_term; 1.  
 DR Pfam; PF00326; peptidase\_S9; 1.  
 SQ SEQUENCE 863 AA; 98263 MW; 40FE0B78E26CED5 CRC64;

Query Match 60.2%; Score 1494.5; DB 4; Length 863;  
 Best Local Similarity 50.5%; Pred. No. 8.5e-121;  
 Matches 281; Conservative 69; Mismatches 105; Indels 101; Gaps 2;

QY 1 TGTANPKVTFKMEIMDAEGRITIDVDKELIQFELLFEGVEYIARAGMTPEGKYMSI 60  
 DB TGSKNPKIALKLAEPQDSQGIKIVSQEKELVQFSSLPFKVEYIARAGMTROCKYAMM 367  
 QY 61 LDRSOTRLQIVLISPELFIPEDDVMERQRLIESVDSVTPLIYEETDIWINHIF 120  
 DB FLDRPQOQLQVLPLPALFIPTENEQRLASARAAYPRNVQPVYVEEVTNWVNVHDI 427  
 QY 121 HVPQSH-EEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFKPIKEI 179  
 DB 428 YPPQAGQDQDFGLRANECKTGCHLYRTVLTAKDQVMTBLSPTBDFKPIKEI 487  
 QY 180 AITSGEWEVLGRGNSIQNDEVRLVFEQTKDPSLEHLLVYVSYNPGVETRLTDGYS 239  
 DB 488 ALTSGEWEVLARHGSKIWNBEETLVYFQGTQTPLEHLLVYVSYEAGSIIVLTGFS 547  
 QY 240 HSCCISOHCDFPISKSNQKNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYP 299  
 DB 548 HSCSMQSQFMFVSHSVSTPCVHYKLSGDDDLHKQPRFMAAMEAACPDPYV 607  
 QY 300 PEIFSESTGTFLYGMLYKPHDLOPKKYPVLFIYIGSPQ----- 340  
 DB 608 PEIFHHTRADVQLYGMIVPHTLQPKRKPTLVFYVGGQVQVNVNSFGIKYLRNLTL 667  
 QY 341 -----VAIAGAPVTLMIFYDTGTYERYMGHPDQNEOGYLGSA 340  
 DB 668 ASLGVAVVVIDGRSGCORGLRFGALKNQMGVEIEDQVGLQFVAEKYFIDLSRYAII 727  
 QY 341 -----VAIAGAPVTLMIFYDTGTYERYMGHPDQNEOGYLGSA 379  
 DB 728 GMSYGFSLMGLIHKQVFKVAIAGAPVTWMAAYDTGYERYMDVPENNQGYEAGSVA 787  
 QY 380 MQAEKPPSEPNRLLLHGFLENVHFHTSILSLFVRACKPYDLOIYPOERSIRVPS 439  
 DB 788 LHVEKLPNEPNRLLLHGFLENVHFHTNPLVSQLIRACKPYOLOIYPERHSIRCPES 847  
 QY 440 GEHYELHLHYLOENL 455  
 DB 848 GEHYEVLTLHFLQEYL 863  
 RESULT 8  
 Q8BMT9 PRELIMINARY; PRT; 862 AA.  
 AC O8BMT9;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Dipeptidyl peptidase 9 homolog.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=liver;  
 RX MEDLINE=22354683; Pubmed=12466851;  
 RA The FANTOM Consortium,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK050021; BAC34034.1;  
 SQ SEQUENCE 862 AA; 98001 MW; BID566E24A834E8 CRC64;

Query Match 60.0%; Score 1488.5; DB 11; Length 862;  
 Best Local Similarity 50.0%; Pred. No. 2.8e-120;  
 Matches 278; Conservative 70; Mismatches 107; Indels 101; Gaps 2;

QY 1 TGTANPKVTFKMEIMDAEGRITIDVDKELIQFELLFEGVEYIARAGMTPEGKYMSI 60  
 DB TGSKNPKIALKLAELQDHOQKIVSSCEKELVQFSSLPFKVEYIARAGMTROCKYAMM 366  
 QY 61 LDRSOTRLQIVLISPELFIPEDDVMERQRLIESVDSVTPLIYEETDIWINHIF 120  
 DB 367 FLDRPQOQLQVLPLPALFIPEVSEKQRAARAAYPRNVQPVYVEEVTNWVNVHDI 426  
 QY 121 HVPQSH-EEIEIFIFASECKTGFRHLKYITSLIKESKYRSSGGLPAPSDFKPIKEI 179  
 DB 427 HPPQAGQDQDFGLRANECKTGCHLYRTVLTAKDQVMTBLSPTBDFKPIKEI 486  
 QY 180 AITSGEWEVLGRGNSIQNDEVRLVFEQTKDPSLEHLLVYVSYNPGVETRLTDGYS 239  
 DB 487 ALTSGEWEVLSRHSKIVNEQRLVYFQGTQTPLEHLLVYVSYEAGSIIVLTGFS 546  
 QY 240 HSCCISOHCDFPISKSNQKNPHCVSLYKLSPEDDPTCKTEFWATILDSAGLPDYP 299  
 DB 547 HSCSMQSQFMFVSHSVSTPCVHYKLSGDDDLHKQPRFMAAMEAACPDPYV 606  
 QY 300 PEIFSESTGTFLYGMLYKPHDLOPKKYPVLFIYIGSPQ----- 340  
 DB 607 PEIFHHTRADVQLYGMIVPHTLQPKRKPTLVFYVGGQVQVNVNSFGIKYLRNLTL 666  
 QY 341 -----VAIAGAPVTLMIFYDTGTYERYMGHPDQNEOGYLGSA 340  
 DB 667 ASLGVAVVVIDGRSGCORGLRFGALKNQMGVEIEDQVGLQFVAEKYFIDLSRYAII 726  
 QY 341 -----VAIAGAPVTLMIFYDTGTYERYMGHPDQNEOGYLGSA 379  
 DB 727 GMSYGFSLMGLIHKQVFKVAIAGAPVTWMAAYDTGYERYMDVPENNQGYEAGSVA 786  
 QY 380 MQAEKPPSEPNRLLLHGFLENVHFHTSILSLFVRACKPYDLOIYPOERSIRVPS 439  
 DB 787 LHVEKLPNEPNRLLLHGFLENVHFHTNPLVSQLIRACKPYOLOIYPERHSIRCPES 846  
 QY 440 GEHYELHLHYLOENL 455  
 DB 847 GEHYEVLTLHFLQEHL 862  
 RESULT 9  
 Q8BVG4 PRELIMINARY; PRT; 862 AA.  
 AC O8BVG4;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Dipeptidyl peptidase 9 homolog.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;  
 RX MEDLINE=22354683; Pubmed=12466851;  
 RA The FANTOM Consortium,  
 RP the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RT Nature 420:563-573(2002).  
 DR EMBL: AK078301; BAC37211.1; -  
 SO SEQUENCE 862 AA; 98049 MW; BAFD3476B6F50030 CRC64;

Query Match 59.6%; Score 1478.5; DB 11; Length 862;  
 Best Local Similarity 49.8%; Pred. No. 2,1e-119;  
 Matches 277; Conservative 70; Mismatches 108; Indels 101; Gaps 2;

QY 1 TGTANPKVTRFMSIMIDAGRIIDVIDKELIQPELIFECVEYIARAGWTPBCKYAWSI 60  
 DB 307 TGSNKPALKLAKLAFQDTHQGIKIVSSCKELVQPPSSLPFVVEYIARAGWTRDGYAWAM 366  
 QY 61 LLDSSQRLQIVLISPELIFVEDDVMERQRLIESVPDVTPLIIEETTDIMINIDIF 120  
 DB 367 FLVYRQRLQIVLRLPALFIPAVESARQAAARAVGNOPFVIEEYTNVWINVADIF 426  
 QY 121 HVPPQSH-EEEIEIFPASECKTGRHLKYTISILKSKYKRSQGLPAPSPFKCPKEEI 179  
 DB 427 HPFQAEQGDQDFCFRANECKTGFCHLRYTVLTKDYDWTETPLSPFDEDFKCPKEEV 486  
 QY 180 AITGGEWVLGRHGSNIQVDEVRILYFEGTKDSFLEHLYVGVYVNGEYTRLTDRGYS 239  
 DB 487 ALITGGEWVLSRHGSKIMVNEOTKLVPQGTDPLEHLYVVSYESGELVRLTLTGFS 546  
 QY 240 HSCCISQHCDFIISKYSNOKNPHCVSLYKLSPPEDDPTCKTEFMATILDSAGPLDYTP 299  
 DB 547 HSCMSQSDFMVFHYHSSVSFPCVHYKLSGPDDELHKQPRFASMEALANCPDVP 606  
 QY 300 PEISFESTGTFTYGMLYKPHDLQPKKYPVTLFYGGPQ----- 340  
 DB 607 PEIHFHTRADVOLYGMLYKPHDLQPKKYPVTLFYVGGPQVOLVNSFKGIKYLRLNTL 666  
 QY 341 ----- 340  
 DB 667 ASLGVAVVVIDGRSSCORGLHFGALKNQMGVEIEDVQGLQYAEKYGITIDLSRAI 726  
 QY 341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEGGYLGSAVA 379  
 DB 727 GMSYGFSLMGLIHKPQVFAIAGAPVTMAVDTGYTERYMDVPENNKQYBAGSAVA 786  
 QY 380 MQAKFPSEPRLLIHGFLDENVFAHTSILSLVYAGKPYDQIYPERHSIRPES 439  
 DB 787 LHVETKLPEPRRLILHGFLEBNVFFHTNPLVSQILRAKRPYDQIYPERHSIRPES 846  
 QY 440 GEHYELHLHYLOENTL 455  
 DB 847 GEHYELHLHYLOENTL 862

## RESULT 10

Q9HBM4 PRELIMINARY; PRT; 310 AA.  
 AC Q9HBM4  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE dipeptidyl peptidase 8 (Fragment).  
 GN DPP8  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=20467194; PubMed=11012666;  
 RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,  
 RA Gorell M.D.;  
 RT "Cloning, expression and chromosomal localization of a novel human  
 RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";  
 RT Eur. J. Biochem. 267:6140-6150(2000).

DR EMBL: AF221635; AAC29767.1; -  
 DR MEROPS; S09.018; -  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 FT NON-TER 1  
 SO SEQUENCE 310 AA; 35396 MW; BE87C34026D9C7AC CRC64;

Query Match 55.5%; Score 1377.5; DB 4; Length 310;  
 Best Local Similarity 83.5%; Pred. No. 2.9e-111;  
 Matches 259; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 207 FEGTKDSPLEHLYVVSYPNGEYTRLTDRGYSHSCCISQHCDFIISKYSNOKNPHCVSL 266  
 DB 1 FEGTKDSPLEHLYVVSYPNGEYTRLTDRGYSHSCCISQHCDFIISKYSNOKNPHCVSL 60  
 QY 267 YKLSPPEDDPTCKTEFMATILDSAGPLDYTPPEISFESTGTFTYGMLYKPHDLQPG 326  
 DB 61 YKLSPPEDDPTCKTEFMATILDSAGPLDYTPPEISFESTGTFTYGMLYKPHDLQPG 120  
 QY 327 KKYPVTLFYGGPQ----- 340  
 DB 121 KKYPVTLFYGGPQGLIEDDQVGLQYLSRYFDLDRVGINMSYGYLSMALMQR 180  
 QY 341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEGGYLGSAVAMQAEKFPSEPRLL 395  
 DB 181 SDIRFVAIAGAPVTLMIFYDTGYTERYMGHPDQNEGGYLGSAVAMQAEKFPSEPRLL 240  
 QY 396 HGFLENVFAHTSILSLVYAGKPYDQIYPERHSIRPESGEHYELHLHYLOENTL 455  
 DB 241 HGFLENVFAHTSILSLVYAGKPYDQIYPERHSIRPESGEHYELHLHYLOENTL 300  
 QY 456 GSRIALAKVI 465  
 DB 301 GSRIALAKVI 310

## RESULT 11

Q9N3F5 PRELIMINARY; PRT; 628 AA.  
 AC Q9N3F5  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN DKFZP762M2413.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Melanoma;  
 RA Ansoorge W., Winkler U., Mewes H.W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB844376; CAD39039.1; -  
 DR InterPro; IPR001375; Peptidase\_S9.  
 DR InterPro; IPR000379; Ser esterase site.  
 DR Pfam; PF00326; Peptidase\_S9; 1.  
 KW Hypothetical protein.  
 FT NON-TER 1  
 SO SEQUENCE 628 AA; 71368 MW; DC086A9440507C3B CRC64;  
 Query Match 51.0%; Score 1266.5; DB 4; Length 628;  
 Best Local Similarity 46.2%; Pred. No. 3.4e-101;  
 Matches 244; Conservative 61; Mismatches 122; Indels 101; Gaps 2;

QY 1 TGTANPKVTRFMSIMIDAGRIIDVIDKELIQPELIFECVEYIARAGWTPBCKYAWSI 60  
 DB 100 TGSNKPALKLAKLAFQDTHQGIKIVSSCKELVQPPSSLPFVVEYIARAGWTRDGYAWAM 159  
 QY 61 LLDSSQRLQIVLISPELIFVEDDVMERQRLIESVPDVTPLIIEETTDIMINIDIF 120  
 DB 160 FLDRPQRLQIVLRLPALFIPSTENEQRLASARAVRNQVPVVEYTNVWINVADIF 219

```

Oy      121 HVEPQSHHE-LEIEFIASQCKTGFRHLUYKITSILKSKYKRSSGGLPAPSEFCPIKEE 179
Db      220 YFFPQORERTSSAFAPMNAAPASAICTYSPPILKSGQIDWSEPFSPGDEDFICPIKEE 279
Oy      180 AITSGEMEWLGRHGSNIQYDEVARLYVEFGTKDSPLEHLLVYVSYNPCGEVTRLTGRYS 239
Db      280 AITSGEMEWLARHGSIKIMYNEEKLYVFOCTORPLEHLLVYVSYEAGIYVLTTPGFS 339
Oy      240 HSCCISQHCDFPISKYSNOKNPHCVSLYKLSPEDDPTCKTEFWATILDSAGPLPDYTP 299
Db      340 HSCSMQSNQDMFEVSHSSVSTPPCVHYVYKLSGDDDBPLHKQPFMSMMMAACSPDYVP 399
Oy      300 PEIIFSESTTGFLYGMILKPHLOQSKKYPTLYFLYCGGQ----- 340
Db      400 PEIIFHTHSQDVRLYGMIKPHALOQSKHPVLYFVGGGQVQLVNNSEFKGIYRLRNLTL 459
Oy      341 ----- 340
Db      460 ASLGYAVVYIDGRGSCQRLRFEGALKNQMGVIEIDQVEGLQFVAEKYFIDLSVAIH 519
Oy      341 -----VAIAGAPVTLMIFYDTGYTERYMGHPDQNEGGYVLAGSVA 379
Db      520 GWSYGGFLSLMGLIHKKPQVFKVAIAAGAPVTVMVAAADTGYTERYMDVPENNQHGXEAGSVA 579
Oy      380 MQAEKFPSEPNRLLILHGFLDENVFAHHSILSLFLVRACKPQYDLOIY 427
Db      580 LAHEKLPNEPNRLLILHGFLDENVFAHHSILSLFLVRACKPQYDLOIY 627

```

```

RESULT 12
ID 075273 PRELIMINARY; PRT; 508 AA.
AC 05.273;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE R26984_1 (Fragment).
OS Homo sapiens (HUMAN).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Tranheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kionmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carraro A.V.;
RT "Sequence analysis of a 2.5 MB region in 19p13.3.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC000594; AAC33801.1; -
DR MEROPS; S09.019; -
DR InterPro; IPR000379; Set_eatse_site.
LT NON TER 1
SQ SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;

```

Query Match	Score	DB	Length
Best Local Similarity	44.4%;	Pred. No. 1.8e-94;	
Matches 241; Conservative	64;	Mismatches 93;	Indels 145;
			Gaps 7;

```

OY      23 I I D V A I D K E I O F F E I I F E B E V Y I A R G W T P E G K A M S I I L D S Q R T L O V I L S P E I F Y  82
      1  I V S T O K E I V O F F S L F P K V E Y I A R G ----- A M A M L D R P Q O M L O V L A P P A L F I S  53
Db
OY      83 E D D V M E R O R L I E S V P S V T P L I I Y E T T D I M I N I H D I F H V F O S H - E E I E I F I A S E C K T  141
      54 T E N E Q R L A S A R A V P N V P Y V V Y E E V T W I N V H D I P A P P O S E E D E I C F L R A N E C K T  113
OY      142 G F R H E Y K I T S I L K E S Y K R K S S G L P A P S D F K C I K E I A I T S G E W E V L R H G S N I O V D E V  201

```

[illegible]

ID	Q9BVR3	PRELIMINARY:	PRT:	439 AA.
AC	Q9BVR3;			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Hypothetical protein (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=placenta;			
RA	Strausberg R.;			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC000970; AAH00970.1; -.			
DR	MEROPS; S09.019; -.			
DR	GeneW; HGNC:18648; DPP9.			
DR	InterPro; IPR001375; Peptidase S9.			
DR	InterPro; IPR000379; Set_eastr_site.			
DR	Pfam; PF00326; Peptidase_S9; 1.			
KM	Hypothetical protein.			
FT	NOR_TIR			
SO	SEQUENCE 439 AA; 49926 MW; A18BA9E12092BAF CRC64;			

Query Match	46.5%	Score 1154.5	DB 4	Length 439
Best Local Similarity	50.6%	Pred. No. 1.1e-91		
Matches 222	Conservative 44	Mismatches 72	Indels 101	Gaps 2

```

QY 118 DIFWVPSH-EEELFFIASECKGFPHLYITSLILKESKTKRSSGGLPARSPDFPCPIK 176
DB 1 DIFPFPSSEGDDELCPFRANECKGFCHLYVTVLVSOGVDWSEPFSGEDEFKCPPIK 60
QY 177 EEAITSGEWEYLGRHSGNIQVDEVRRLVYEEGTQDSPLEHILYVSYNPGSEVTRLTDR 236
DB 61 EEAILTSEWEYLARHSGKIWNBEETKLIVFGQTDTPLEHILYVSYSTAAAEIYRLTTP 120
QY 237 GYHSCTISQHCDFIISKYSNQKNPCHVSLYKLSSPEDDPCTCKYEFWATILDSAGPLPD 296

```

```

Db      121 GFSSCSMSQNFPMFVSHYSSVSTPPCVHYVYKLSGPDDBLHKQPRFWMASMEASCPDP 180
Qy      297 YTPPEIFSFESTGFTLYGMLYKPHDLPQPKKYPTVLFYIGPQ----- 340
Db      181 YVPEIFPHFTTRSDVRLYKMYKPHALQPKKHFTVLFYVGQVQVLVNNSPFKIKYLR 240
Qy      341 ----- 340
Db      241 NTLASLGAVVVIDGRSGCRGLRFEGALKNQMGQVEIEDQVEGLQFAEKYGFIDLSRV 300
Qy      341 -----VAIAGA PVTLMITFYDGYTERVYMGHGDQNEQGYLLG 376
Db      301 AIHGWSYGLSLMGLIHKQVPEKATAGAPVTVMAYDGYTERVYMDVENNQHGVEAG 360
Qy      377 SVAMQAEKFPSEPNRLLLHGFLENVHFHTSILSLFVRAKGPYDLOQYPOERHSIRV 436
Db      361 SVALLHVEKLPNERRLILHGFLENVHFHTNPLVSQLIRAKGPYDLOQYPOERHSIR 420
Qy      437 PESGEHYELHLHYLOENL 455
Db      421 PESGEHYEVTLLHFLQEYL 439

```

## RESULT 14

```

ID      09HBM2      PRELIMINARY;      PRT;      360 AA.
AC      09HBM2;
DT      01-MAR-2001 (TEMBLrel. 16, Created)
DT      01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT      01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE      Dipeptidyl peptidase 8 (Fragment).
GN      DPP8.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_TaxID=9606;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20467194; PubMed=11012666;
RA      Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA      Gorrell M.D.;
RT      "Cloning, expression and chromosomal localization of a novel human
RL      dipeptidyl peptidase (DPP) IV homolog, DPP8."
EMBL; AF221637; AAC29769.1; -.
DR      MEROPS; S09.018; -.
DR      InterPro; IPR001375; Peptidase_S9.
DR      InterPro; IPR000379; Ser_estrs_site.
DR      Pfam; PF00326; Peptidase_S9; 1.
FT      NON_TER      1
FT      TER      360
SQ      SEQUENCE      360 AA; 41070 MW; CF81COBB61423E22 CRC64;

```

```

Query Match      34.7%; Score 860.5; DB 4; Length 360;
Best Local Similarity 40.5%; Pred. No. 2.8e-66;
Matches 186; Conservative 0; Mismatches 0; Indels 273; Gaps 2;
Qy      1 TGTANPKVTFKMEIMIDAGRIIVDKELIQPFILFEGVEYIARAGWTPGKYAWSI 60
Db      75 TGTANPKVTFKMEIMIDAGRII----- 98
Qy      61 LLDRSQTRLOIVLISELPFVDDDWERQRLIESVDSVTLPIIYEETTDININIHDF 120
Db      99 ----- 98
Qy      121 HVEPQSHHEIEIFIFASECKTGFRLHYKTSILKSKYKRSGLPAPSPKPIKEEIA 180
Db      99 ----- 98
Qy      181 ITSGEMEVLYGRHSGNIQVDEVRLLVYFEGTKDSPLEHHLVYVSYNPGEVTRLTRDGRYSH 240
Db      99 -----VDEVRLLVYFEGTKDSPLEHHLVYVSYNPGEVTRLTRDGRYSH 141

```

```

Qy      241 SCGISQHCDFEISKYSNKNKPHCVSLYKLSPPEDDPTCKTKFWMATILDSAGLPDYTP 300
Db      142 SCGISQHCDFEISKYSNKNKPHCVSLYKLSPPEDDPTCKTKFWMATILDSAGLPDYTP 201
Qy      301 EIFSPESTGFTLYGMLYKPHDLPQPKKYPTVLFYIGPQ----- 340
Db      202 EIFSPESTGFTLYGMLYKPHDLPQPKKYPTVLFYIGPQVQVLVNNRFGKVKYFRNLTLA 261
Qy      341 ----- 340
Db      262 SLGIVVVVIDNRSGCHRGKLFEGAFKYNQGLEIDDOVEGLQVLASRYDILDDRVGIGH 321
Qy      341 -----VAIAGA PVTLMITFYDGYT 359
Db      322 WSYGYLSLMAIMQSRDIFRVALAGAPVTLMITFYDGYT 360

```

## RESULT 15

```

ID      09VC20      PRELIMINARY;      PRT;      1042 AA.
AC      09VC20;
DT      01-MAY-2000 (TEMBLrel. 13, Created)
DT      01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT      01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE      CG3744 protein.
GN      CG3744.
OS      Drosophila melanogaster (fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
NCBI_TaxID=7227;
OC      Ephydroidea; Drosophilidae; Drosophila.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=BERKELEY;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
RA      Borkovetz D., Borchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA      Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Fouts C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA      Jalali M., Kalish F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merklou G., Milshina N.V., Mobarry C., Morris J., Mosher J.A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA      Palacios M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasarman D.A., Weisbrock G.M., Weissbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:15:46 (Search time 5076.73 seconds  
(without alignments)  
3747.092 Million cell updates/sec

Title: US-10-070-464-5  
Sequence: 1 TGTATPKYTFKMSIIMDAE.....HLHLHYQENIGSRIALKVI 465  
Perfect score: 2482

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO/spool/US10070464/runat.15102003\_113553\_24810/app.query.fasta\_1.2652  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=otc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10070464 @CGN 1.1 13561 @runat.15102003\_113553\_24810 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEUDERY -NEG SCORES=50 -WAIT -DSPBLOCK=100 -IONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*

1:	gb_ha:*
2:	gb_hcg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pac:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_scs:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_va:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pac:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_scs:*
28:	em_un:*

29: em\_vi:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pin:\*

35: em\_hcg\_rod:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrt:\*

38: em\_sy:\*

39: em\_hcgo\_hum:\*

40: em\_hcgo\_mus:\*

41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	1669	9 AF221636	AF221636 Homo sapi
2	2482	100.0	2668	6 AX405771	AX405771 Sequence
3	2482	100.0	2642	6 AX405770	AX405770 Sequence
4	2485	98.9	2778	9 AK000290	AK000290 Homo sapi
5	2447.5	98.6	3130	9 BC030688	BC030688 Homo sapi
6	2446.5	98.6	4523	6 AX608731	AX608731 Sequence
7	2441.5	98.4	2161	6 BD157001	BD157001 Primer fo
8	2441.5	98.4	2161	9 AK027826	AK027826 Homo sapi
9	2432	97.6	2649	6 AX354795	AX354795 Sequence
10	2432	97.6	2656	6 AY172659	AY172659 Sequence
11	2422	97.6	2671	6 AX608725	AX608725 Sequence
12	2422	97.6	3106	6 AX342633	AX342633 Sequence
13	2422	97.6	3127	9 AF221634	AF221634 Homo sapi
14	2422	97.6	3143	6 AX354793	AX354793 Sequence
15	2422	97.6	4535	6 BC040203	BC040203 Homo sapi
16	2418.5	97.4	4476	6 AX608743	AX608743 Sequence
17	2402	96.8	4829	6 AX608735	AX608735 Sequence
18	2362	93.7	4799	10 BC043124	BC043124 Mus muscu
19	2244.5	90.4	2510	6 AX338497	AX338497 Sequence
20	2149	86.6	4685	6 AX608745	AX608745 Sequence
21	1494.5	60.2	2602	9 AY172660	AY172660 Homo sapi
22	1494.5	60.2	2617	6 AX608727	AX608727 Sequence
23	1494.5	60.2	3000	9 AF542510	AF542510 Homo sapi
24	1494.5	60.2	3716	6 AX480934	AX480934 Sequence
25	1494.5	60.2	4219	6 AX608751	AX608751 Sequence
26	1494.5	60.2	4232	9 BC037948	BC037948 Homo sapi
27	1494.5	60.2	4295	9 AF452102	AF452102 Homo sapi
28	1494.5	60.2	4302	6 AX608747	AX608747 Sequence
29	1491.5	60.1	2676	6 AX524928	AX524928 Sequence
30	1445	58.2	4180	6 AX608759	AX608759 Sequence
31	1445	58.2	4263	6 AX608757	AX608757 Sequence
32	1386	55.8	4309	6 AX608737	AX608737 Sequence
33	1377.5	55.5	1197	9 AF221635	AF221635 Homo sapi
34	1342.5	54.1	4076	6 AX608755	AX608755 Sequence
35	1342.5	54.1	4159	6 AX608753	AX608753 Sequence
36	1338	53.9	3243	9 HSM805448	AL834376 Homo sapi
37	1293	52.1	4037	6 AX608763	AX608763 Sequence
38	1293	52.1	4120	6 AX608761	AX608761 Sequence
39	1274.5	51.3	2261	6 AX713363	AX713363 Sequence
40	1274.5	51.3	2261	9 AK054656	AK054656 Homo sapi
41	1154.5	46.5	2546	9 BC000970	BC000970 Homo sapi
42	1055	42.5	753	6 AX524935	AX524935 Sequence
43	860.5	33.7	1083	9 AF221637	AF221637 Homo sapi
44	841.5	33.9	3740	3 BT001499	BT001499 Drosophila
45	673.5	27.1	215711	3 AC008205	AC008205 Drosophila

RESULT 1 0 4 ALIGNMENTS



AF221636  
 LOCUS AF221636 1669 bp mRNA linear PRI 05-NOV-2000  
 DEFINITION Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds,  
 alternatively spliced.  
 ACCESSION AF221636  
 VERSION AF221636.1 GI:11095191  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1669)  
 AUTHORS Abbot, C.A., Yu, D.M., Woollett, E., Sutherland, G.R., McCaughan, G.W.,  
 and Gorrell, M.D.  
 TITLE Cloning, expression and chromosomal localization of a novel human  
 dipeptidyl peptidase (DPP) IV homolog, DPP8  
 JOURNAL Eur. J. Biochem. 267 (20), 6140-6150 (2000)  
 MEDLINE 20467194  
 PUBMED 11012666  
 REFERENCE 2 (bases 1 to 1669)  
 AUTHORS Abbot, C.A., Yu, D., McCaughan, G.W. and Gorrell, M.D.  
 TITLE Direct Submision  
 JOURNAL Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver  
 Centre, Centenary Institute of Cell Biology and Cancer Medicine,  
 Locked Bag No. 6, Newtown, Sydney, NSW 2042, Australia  
 Location/Qualifiers  
 1. 1669  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /map="15q22"  
 /chromosome="15"  
 /clone="T8"  
 /clone\_type="T8"  
 /clone\_type="placenta"  
 /clone\_lib="lambdaTriplexTM: Clontech"  
 <1. 1669  
 /gene="DPP8"  
 /gene="DPP8"  
 <1. 1399  
 /note="serine protease; member of prolyl oligopeptidase S9b  
 family belonging to enzyme clan SC; isoform missing the  
 catalytic serine required for enzyme activity; additional  
 splice variants of this mRNA are found in Genbank  
 Accession Numbers AF221634, AF221635 and AF221637"  
 /codon\_start=2  
 /product="dipeptidyl peptidase 8"  
 /protein\_id="AAG29768.1"  
 /db\_xref="GI:11095192"  
 /translation="TGTANPKVTFPKMSEIMIDAEGRIDVIDKELIQPELIEGVEY  
 IARAGTPEGKYAMSLIDRSQRLQIVLISPELTPVEDDVMERQLIESVDSVTP  
 LIIVETTDIWINIHDIHFVPOSHHEEIEPTFASCKTGFPHLYTISILKSKYKR  
 SSGGAPAPDFKPCIKKEIJAITSGEWEVNSIQOVENRLVFEETKDSPEHH  
 LYVSVYNGEVTRLTRDGRVSGCCISQHCDFISIKYSQKPKHCISLYKLSPEDD  
 TCKTKEFATIIIDSGAPLDPDTPPELRSFESTGTFTLYGMLYKPHDLOGKRYTLYL  
 IYGGPOVAIAGAVTIDWITFYDTGYTERIKGHPDQNGYITGDSVAAQAKFSEPRLL  
 ILLHGLDENVHFAHTSILLSFLVRAGKPYDLQIYQERHSTRVPSGEHYELHLY  
 LOENLSRIALAKVI"  
 BASE COUNT 524 a 336 c 336 g 473 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,466-236 Length: 1669  
 Score: 2482.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-070-464-5 (1-465) x AF221636 (1-1669)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPhelyMetSerGluIleMetIleAspAlaGlu 20  
 DB 2 ACAGGTACAGCAATCTTAAGTCACTTTTAAGATGTCAGAAATATATGATTGATGCTGA 61

QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleuPheGlu 40  
 DB 62 GGAAGATCATAGATGTCATAGATTAAGAACTAATTCACCTTTGAGATTCTATTGA 121  
 QY 41 GlyValGluTyrIleIleAargAlaGlyTyrThrProGluGlyLysTyrAlaTTPSerIle 60  
 DB 122 GGAGTTGAATATATTATCCAGAGCTGATGATGATCTCCAGAGAAATATATCTGCTCANC 181  
 QY 61 LeuLeuAspArgSerGluThrArgLeuGluIleValIleuIleSerProGluLeuPheIle 80  
 DB 182 CTACTAGATCGCTCCAGACTGCTGCTACAGATAGTGTGATCTCACCTGAATTAATTATC 241  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 242 CAGATGAAGATGATGTTATGGAAGAGAGACATCATGATGATGATGCTGATCTG 301  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 302 ACGCCACTAATTATCTATGAGAAACAACAGACATCTGGATTAATATCATGACATCTTT 361  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 362 CATGTTTTCCTCCAAAGTCACAGAGAGAAATTGAGTTATTGCTCTGAATGCAA 421  
 QY 141 ThrGlyPheArgHisLeuTyrIleTyrSerIleLeuLeuGluSerIleTyrLysArg 160  
 DB 422 ACAGGTTCCTCCATTTATACAAATATCATCTATTTTAAAGAAAGCAATATTAACA 481  
 QY 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 DB 482 TCCAGTGGGCTGCTGCTGCTCCCAAGTATTCAGTGCCTATCAAAAGAGATAGCA 541  
 QY 181 IleThrSerGlyGluThrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 542 ATTACAGGTGGATGGAGAGTTCGGCCGCAATGATTAATATCAAGTATGATGA 601  
 QY 201 ValArgArgLeuValTyrPheGluGluTyrThrLysAspSerProLeuGluIleHisLeuTyr 220  
 DB 602 GTACAGAGCTGTATATTTTAAGGACCAAAAGACTCCCTTTAGAGATCACCTGTAC 661  
 QY 221 ValValSerTyrValAsnProGluGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 662 GTAGTCAGTACGTAAATCTCTGAGAGGAGCAAGGCTGACGACCTGCTACTACAT 721  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 722 TCTTGCTGATCAGTCAGCATGTACTTCTTTATAGTAAGTATAGTAACAGAAAGAT 781  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 782 CCACACTGTGTCTCCCTTTACAAAGCATCAATCTCGAAAGATGACCCAACTTGCAAAACA 841  
 QY 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 DB 842 AAGGAATTTGGGCCACCACTTTTGATTCAGAGAGCTCTCCCTCGACATATCTCTCCA 901  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 902 GAAATTTCTCTTTGAAAGTACTACTGATTTAATGTATGAGATGCTCTACAGACT 961  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 962 CATGATCTACAGCTCGAAAGAAATATCTACTGCTGCTTATATATGATGCTGCTCAG 1021  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrPheTyrAspThrGlyTyrThrGlu 360  
 DB 1022 GTTGTATTGCTGGGACCCCACTCACTGTGATCTCTATGATACAGATACAGGAA 1081  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 1082 CATTATATGGGTACACCTGACCAAGATGAACAGGCTATTACTTACGATCTGTGCGCATG 1141

QY 381 GlnAlaGluPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
DB 1142 CAAGCAGAAAAGTCCCTCTGACCAATCGTTACTCTCTTACATGCGTTCCGGAT 1201  
QY 401 GlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 1202 GAGATGTCATTTTSCACATACACAGTATATTACTAGATTTTCTAGAGGGCTGGAAAG 1261  
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
DB 1262 CCATATGATTTACAGATCTTACTCTCAGAGAGCAGACAGATAGAGCTTCTGATGGGA 1321  
QY 441 GlnHisTyrGluLeuHisIleLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
DB 1322 GAACATTAAGAACTGATCTTTTGACACTACCTTCAAGAAAACCTTGATCAGCTATTGCT 1381  
QY 461 AlaLeuLysValIle 465  
DB 1382 GCTCTAAAGTGATA 1396  
RESULT 2  
AX405771 2668 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 186 from Patent WO0222660.  
DEFINITION AX405771  
ACCESSION AX405771  
VERSION AX405771.1 GI:21438981  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,  
Xue, A.D., Yang, Y., Weinman, T., and Drmanac, R.T.  
TITLE Novel nucleic acids and polypeptides  
JOURNAL Patent: WO 0222660-A 186 21-MAR-2002;  
HYSEQ, INC. (US)  
FEATURES  
source location/Qualifiers  
1..2668  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
234..2408  
/codon\_start=1  
/protein\_id="CAD34878.1"  
/translation="MAAAMEEQQLGVEIPEADCEBNTESQDRPKLEPFYVERYSMQ  
LKLLADTRKHYGMMAKAPHDPMFVKRNDPGPSDRIRYDAMGEMENLTFRSEI  
PTIRAAVLAMSMKPLDLFOOQPLRPVLVETSCPNIRMDKLCPADDMWAFIHSN  
DIWISNITVREERLTLYVNEELANMEEDRSAGVAFVQEFDRYSGWMPKAEET  
DSGGKILRLYFENESEVEITHTVSPMLETRADSFRIKGTANPKVTFPMSEIMI  
DAEGRIIDVIDELIQPEILFEGVEYIARAWTPGKAWMSILLDRSGTRLIQILIS  
PELFIPEVDVVERQRLISVDSVTPLIIVETTDIWINHIDIFHVFQSHBEIEIF  
IFASECKTFRLYKITSLIKESKYKSGGLPARSDFCPIKEELAIKSGMEVILGR  
HGSNIQVDEVRLLYFEGTKDSPLDHLIYVSVANGVETRLIDGYSHSCCISQHD  
FRISKYSNQKNPHCVSLYKLSPEDDPTCKTKPEWATILDSAGRPIDTPPTPTSES  
TTGFTLGMILYKPHDQPKKYPIVLFTEGQVAVIAGAPVLMIFDYGTERTWGH  
PDQNGYLLGVANAOEKRPSEPNRLHLHGFIDENVHFAHTSILSLFLVAGKPYD  
LQIYPOERHSISVPESGEHYELHLHYLOENIGSRIALKVI"

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 1011 AAGGATACGCAAACTCTAAAGTCACTTTAAAGATCTCAAGAAATATATGATGCTGAA 1070  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1071 GGAAGGATCATAGATGTCATATGATATAGAACTAAATTCAACCTTTTGAGATTCTATTGAA 1130  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTTPThrProGluGlyLysTyrAlaTyrSerIle 60  
DB 1131 GAGATTGAATATATATTCAGACCTGATGACTCTCGAAGGAAATATCTTGCTCATC 1190  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1191 CTACTAGATGCTCCAGACTGCTCTACAGATAGTGTATCTCACCTGAATATTATTATC 1250  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1251 CCAAGTAGAAGATGATGTTATGGAAGGACAGACATTCATGAGTCAAGTCTGATCTGTG 1310  
QY 101 ThrProLeuIleIleIleTyrGluGluThrThrAspIleTTPIleAsnIleHisAspIlePhe 120  
DB 1311 ACGCACTAATTATCTATGAGAAACACACACATCTGATATTAATCATGACATCTTT 1370  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1371 CATGTTTTCCTCCAAAGTACAGAGAGAAATGAGTTTATTTTCCCTGAATCCAA 1430  
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
DB 1431 ACAGGTTTCGGCATTTATACAAATATACATGTTTAAAGAAACAAATATTAACGA 1490  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
DB 1491 TCCAGTGGGGCTGCTGCTGCTCCCAAGTATTCAGTGTCCATCAAGAGAGAGAGA 1550  
QY 181 IleThrSerGlyGluTrogIuValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1551 ATTTACAGTGGGAAATGGAAAGTCTTGCCGCGCATGGATCTTAATATCCAGATTGATGA 1610  
QY 201 ValArgArgLeuValTyrPheGluGlyLysThrLysAspSerProLeuGluHisHisLeuTyr 220  
DB 1611 GTCCAGAGGCTGTATATTTTGAAGGACCAAGAACTCCCTTTAGAGCATCACCTGTAC 1670  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1671 GTAGTCAGTACGTAAATCTCGAGAGGTGACAAAGCTGACCGTGTACTACAT 1730  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
DB 1731 TCTTCTGATCAGTACGACATGCTGCTCTTTATAGTATAGTATATGAAAGAAAT 1790  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1791 CCACACTGTGTCTCTTTACAAGCTATCAATCTCGAAGATGACCAACTTGCAAAACA 1850  
QY 281 LysGluPheTTPAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 1851 AAGGATTTTGGGCCACATTTTGATTCAGACGCTCTCTTCGACTATACTCTCCA 1910  
QY 301 GlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 1911 GAAATTTCTCTTTGAAAGTACTAGATTTTACATGATGAGATGCTTCAAGCCT 1970  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
DB 1971 CATGATCTACAGCCCGAAAGAAATATCTACTGTGCTGTTCATATATGCTGCTCAG 2030  
QY 341 ValAlaIleAlaGlyValProValThrLeuTTPIlePheTyrAspThrGlyTyrThrGlu 360  
DB 2031 GTTGCATTTGCTGGGCCCCAGCTCACTGTGATCTTCTATGATACAGGATACAGGAA 2090  
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380

US-10-070-464-5 (1-465) X AX405771 (1-2668)

DB 2091 CGTTATATGGCTCACCCCTGACCGAATGAAACAGGCTATTACTAGATCTGGCCATG 2150  
QY 381 GlnAlaGluysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
DB 2151 CAGCGCAAAAAGTTCCCTCTGACCAATCGTTTCTGCTTACAGGTTTCCGAGAT 2210  
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 2211 GGAATGTCATTTTCCACATACCAGATATTACTAGGTTTCTTACGAGGCTGGAAAG 2270  
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
DB 2271 CCATATGATTTTACAGATCTATCTCCAGAGAGACACAGCATTAAGACTCTGATCGGGA 2330  
QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
DB 2331 GAACATTAATGAACTGCACTCTTTGCACTACCTTCAAGAAAACCTTGATCATCTATTGCT 2390  
QY 461 AlaLeuLysValIle 465  
DB 2391 GCTCTAAAGTGATA 2405

## RESULT 3

AX405770 2842 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 185 from Patent WO0222660.  
ACCESSION AX405770  
VERSION AX405770.1 GI:21438979  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1  
Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,  
Xue, A.U., Yang, Y., Wehrman, T. and Dymnac, R.T.  
TITLE Novel nucleic acids and polypeptides  
JOURNAL Patent: WO 0222660-A 185 21-MAR-2002;  
HYSEQ, INC. (US)

## FEATURES

source 1..2842  
Location/Qualifiers  
1..2842  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
234..2582  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD34877.1"  
/db\_xref="GI:21438980"

translation="MAAAETEQLGVEIFETADCEENIESODRPKLEPFYVERYSWQ  
LKKLADTRKYHGYMAKAPHDPMFVKRNDPDGPHSDRIYVLAAMGENENENLIFYSEI  
PKTINRAAVLMSMKPLDLFOATLDYGVYSREELERKRIIGVAGISYHOGSG  
TFLFOGSGIYHKDGGPOGTOOPLRPLMVEISCPNIMODPKCADDDMTAFIHSN  
DIWISNIVTRERRLTYVANELANMEEDRASGAVTFVLOEFPDYSGWMPKPKAET  
BSGKIRLILYEENDESEVEIITHVSPMLFETRADSFRIYPTGTANPKVPMSEIMI  
DEGRIDIVIDELIQPEFILLEGEVYIARAGWTPGKAWMILDRSQRILQVILS  
BELPIPEVDVMEQRRLIESVDSVPLIIVETDIDMINIDIFHVFSQHEEIER  
IPASCKTGFRLYKITSILKESKYRSGSGLPASDFKCPIKEELIATSGMEVIGR  
HGSNIQVDEVRLLVEEGTKDSEPLLEHLVVSVYNGEVTRILDRGYSHSCCISOHC  
FISKYSNOKNHCVSILYKLSPEDDPCTKTEFWATILDSAGPLPDYRPIRIFSES  
TTGPTUKMLYKPHDLQPKKYPTULFIYGQVAILACAPVLMIFDYGTERTKMG  
PQNEGGTILGSVAQAEKRPSPNRLHLHGFLDSNVHFHTSILSLFLVABKPYD  
LDIYPOERSISIVPESGEHYELHLHYLOENLSRIALAKVI"

BASE COUNT 857 a 592 c 635 g 758 t

## Alignment Scores:

Pred. No.: 3 03e-236 Length: 2842  
Score: 2482.00 Matches: 465  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0  
US-10-070-464-5 (1-465) x AX405770 (1-2842)  
QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 1185 ACAGGTACAGCAAACTCCAAAGCTCACTTTAAGATCTCAGAAATATATGATGCTCGAA 1244  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1245 GGAAGATCATAGATGCTATAGATAGGAACCTAATTCACCTTTGAGATTCATTGGA 1304  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyrAlaTyrSerIle 60  
DB 1305 GGAAGTTGAATATATGTCAGACCTGGAATGATCTCCGAGAGAAAATATCTTGCTCATC 1364  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1365 CTAAGTGGATCGCTCCAGACCTGCTACAGATAGTGTGATCTCACTGAATTAATTATC 1424  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1425 CCAGTACAGATGATGCTTATGGAAGGACAGACATCATTGATGATGCTGATTTCTGAG 1484  
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
DB 1485 ACGCCACTAATTTATCTATGAAAGAACACAGACATCTGGATTAATTCATGACATCTTT 1544  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1545 CATGTTTTCCCAAGTCACAGAGAGAAATGAGTTATTTTGGCTGATGATCCAAA 1604  
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
DB 1605 ACGGTTCCGTCATTTATACAAAATTTACATCTATTTTAAAGAAACAAATTAACCA 1664  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
DB 1665 TCCAGTGGGAGGCTGCTGCTCCAGATGATTTCAAGTGCTCTATCAAGAGAGATAGCA 1724  
QY 181 IleThrSerGlyGluThrGluValLeuGluValArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1725 ATTACCAAGTGGATGAGGAACTTCTTGCCCGCATGATCTTAATATCCAAAGTATGAGA 1784  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyr 220  
DB 1785 GTCAGAGGCTGTATATTTTBAAGCACCAAGACTCCCTTTAGAGATTAACCTGTAC 1844  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1845 GTAGTCAGTTACGTAATCTCTGAGAGGAGCAAGGCTGACCGTGCTACTACAT 1904  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnIleLysAsn 260  
DB 1905 TTTCTGATAGTCAGACGATGATCTTTATTAAGTAATTAAGTAACCAAGAAAT 1964  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1965 CCACACTGCTGCTCCCTTTACAAAGCTATCAAGCTCTGGAAGATGACCAACTGCAAAACA 2024  
QY 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 2025 AAGGAATTTTGGCCACCATTTTGGATTACAGAGGCTCTCTGACATATCTCTCCA 2084  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 2085 GAAATTTTCTCTTTGAAAGTACTGATTTTACATGATGGAAGTGTCTACAGGCT 2144  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
DB 2145 CATGATCTACAGCTCGAAGAAATATCTCTGCTGCTGCTCATATATGATGCTGCTCAG 2204  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360

|||||  
Db 2205 GTTCTCTATGCTGGGGCCCGCATCTCTGTGATCTTCTATGATACAGATACACGGAA 2264  
361 ArgyrhetesGlyH1sProaspGlnasnGluInglYrYrLeuGlySerValAlaMet 380  
2265 CGTATATGGGTACCCCTGACAGATGAACAGGCTATTACTTATGAGATCTGTGACCATG 2334  
381 GlnaIagluYrPheProSerGluProAsnArgLeuLeuLeuLeuH1sGlyPheLeuAsp 400  
2225 CAACGAGAAAAGTCCCTCTGAAACCAATGCTTACGCTTACAGATGTTCTCTGAT 2384  
401 GlnAsnValHisPheAlaHisThrSerLeuLeuSerPheLeuValArgAlaGlyAs 420  
2385 GAGATATGCCATTTTGGACATACAGATATATTAATGATTTTATGAGGCGCTGAAAG 2444  
421 ProYrAspLeuGlnIleYrProGlnGluArgHisSerIleArgValProGlnSerGly 440  
2445 CCATATGATTTTACAGATCTATCTCAGAGAGACACAGCATTAAGATTCCTGAATCCGGA 2504  
441 GlnHisYrGluLeuH1sLeuLeuH1sYrLeuGlnGluAsnLeuGlySerArgIleAla 460  
2505 GAACATTATGACATGCACTTTTGCACACTTCAAGAAAACCTTGATCATCGTATTGCT 2564  
461 AlaLeuYsValIle 465  
2565 GCTCTAAAGTGATA 2579  
RESULT 4  
AK000290 2778 bp mRNA linear PRI 22-FEB-2000  
LOCUS Homo sapiens cDNA FLJ20283 f1s, clone HEP04088.  
DEFINITION AK000290.1 GI:7020272  
VERSION oligo capping; f1s (full insert sequence).  
KEYWORDS Homo sapiens (human)  
SOURCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
ORGANISM  
REFERENCE  
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,  
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,  
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,  
Nakamura,Y., Isegai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 2778)  
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isegai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
COMMENT  
TITLE Direct Submission  
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,  
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,  
Tel:81-3-5449-5386, Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction; 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source  
1..2778  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP04088"  
/cell\_line="HePG2"  
/cell\_type="hepatoma"  
/clone\_lib="HBP"  
/note="cloning vector pME18FLJ3"  
163..1572  
/note="unnamed protein product"  
/codon\_start=1

/protein\_id="BA091059.1"  
/db\_xref="GI:7020273"  
/translation="MAAAETEOLGVEIFETADCEENIESDPRKLEPFYVERYSWSO  
LKULNATRRKTHGMAKAPHDPIFVRANDPOGHSRIITYLANSGERENTLRYSET  
KPTINRAVLNLSKPLDLFOATLDGMSREBELRERKICTGVIGASTDYHQSG  
TFLEAGSGIYHVWDGPGQFTQOPLPNLVETSPVIRMDPKLCPADPMIAFISHN  
DIWISNVTREERRLTYVHNELANMEDASAGVATLOEPEPRYSGYWCPEAETT  
PSGKILRIIYEENDESEVEIIVTSPMLERRRDSRYRKYRTPKRVKMEIMI  
DAGRIITDVIDKELIOPEFELFEGVEYIARAGWRECKVAMSILLDSORLOIVLS  
PELPIPEDDVMERQRLIESVPDSVTPLLIYEETDIIWNIHDFHVPQSHSEETEF  
IFASECTGFRHLYKITSI"  
BASE COUNT 856 a 562 c 615 g 745 t  
ORIGIN  
Alignment Scores:  
Pred. NO.: 1,41e-233 Length: 2778  
Score: 2455.00 Matches: 464  
Percent Similarity: 99.78% Conservative: 0  
Best Local Similarity: 99.78% Mismatch: 1  
Query Match: 98.91% Indels: 1  
DB: 9 Gaps: 0  
US-10-070-464-5 (1-465) x AK000290 (1-2778)  
QY 1 ThrGlyThraAlaasnProlysValThrPheLyseSerGluIleMetIleAspAlaGlu 20  
Db 1114 ACAGTACACCAAAATCTTAAGTACACTTTTAAAGTGCAGAAATATGATGATGCTGAA 1173  
QY 21 GlyArgIleIleAspValIleAspYsGluLeuIleGlnProPheGluIleLeuPheGlu 40  
Db 1174 GGAAGCATCATATGATGATCATATGAAAGAACTAATTCACCTTTGATGATTTGAA 1233  
QY 41 GlyValGluYrIleAlaArgAlaGlyTrpThrProGluGlyYsYrAlaTrpSerIle 60  
Db 1234 GGAAGTGAATATATTTGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1293  
QY 61 LeuLeuAspArgSerGlnTrpArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1294 CTACTAGATGCTGCCAGACTCCGCTACAGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1353  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
Db 1354 CCAGTATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1413  
QY 101 ThrProLeuIleIleYrGluGluTrpThrAspIleTrpIleAsnIleHisAspIlePhe 120  
Db 1414 ACCGCACTAATTTATATGAAAGAAACACAGACTGATGATGATGATGATGATGATGATGAT 1473  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysVal 140  
Db 1484 CATGTTTTCCTCCAAAGTCCAGAAAGCAAAATGAGTTTATTTTGGCTCTGAAATGCAAA 1533  
QY 141 ThrGlyPheArgHisLeuYrYsIleThrSerIleLeuYsGluSerYsYrYsArg 160  
Db 1534 ACAAGTTTCGTCATTTATATCAAAATTAACATCAT-TTAAAGGAAAGCAAAATTAACGA 1592  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheYsCysProIleYsGluIleAla 180  
Db 1593 TCCAGTGGTGGCTGCTGCTCCAGAGTATTTCAAGTGTCTTCAAAAGAGGATGACA 1652  
QY 181 IleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
Db 1653 ATTAACAGTGTGATATGGAGAGTCTTGGCGGCAAGATGATTAATATCAAGATGATGAA 1712  
QY 201 ValArgArgLeuValYrPheGluGlyThrYsAspSerProLeuGluHisHisLeuYr 220  
Db 1713 GTCAAGAGGCTGTATATTTTGAAGGACCAAAAGATCCCTTTAAGCATCATCCTGAC 1772  
QY 221 ValValSerYrValAsnProGlyGluValThrArgLeuThrAspArgGlyYrSerHis 240  
Db 1773 GTAGTCAGTTACCTTAATCTCGAGAGGTGACAGGCTGACTGACCGTGGCTACTCAT 1832  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerYsYrSerAsnGlnYsAsn 260

```

Db      1833  TCTTGGTCGACATCAGACACACTGACCTTTTATTAAGTATGATATACACAGAGAT 1892
Qy      261  ProHisCysValSerLeuTyrIleuSerSerProGluAspAspProThrCysIleuThr 280
Db      1893  CCACACTGTGTCTCCCTTACAGCTATCAAGTCTGAGATGACCCAACTTGGAAACA 1952
Qy      281  LysGluPheTyrPalaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 300
Db      1953  AAGGAATTTTGGGACCATTTTGGATTGAGAGAGTCTCTTCCGACTATATCTCTCCA 2012
Qy      301  GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIlePro 320
Db      2013  GAAATTTTCTCTTTGAAAGTACTAGATTTTACATTGATGAGATCTCTACAGGCT 2072
Qy      321  HisAspLeuGlnProGlyIleLeuTyrProThrValLeuPheIleTyrGlyProGln 340
Db      2073  CATGATCTACAGCCCTGGAAGAAATATCTACTGCTGCTCATATATGCTGCTCTAG 2132
Qy      341  ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 360
Db      2133  GTTGGTATTTGCTGGGGCCCACTGCTGATCTCTATGATACAGATACAGGAA 2192
Qy      361  ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrIleuGlySerValAla 380
Db      2193  CGTATATAGGTCACCTGACCAAGATACAGGAGCTATTAAGATCTGTGGCCATG 2252
Qy      381  GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuIleGlyPheLeuAsp 400
Db      2253  CAACGAGAAAGTTCCCTCTGACCAATCGTTTACTGCTTATCATGCTTCTGAT 2312
Qy      401  GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 420
Db      2313  GAGATATGCCATTTTGGACATACCGATATATTATGAGATTTTATGAGCGCTGAAAG 2372
Qy      421  ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 440
Db      2373  CCAATATGATTTTACAGATTTATCTCCAGAGAGACACACATTAAGGTTCCGAAAT 2432
Qy      441  GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIle 460
Db      2433  GAACATTATGAACGTCATCTTTTGGACTACCTTCMAAGAAACCTTGATCATCTAT 2492
Qy      461  AlaLeuLysValIle 465
Db      2493  GCTCTAAAGTGATA 2507

RESULT 5      3130 bp  mRNA  linear  PRI 30-MAY-2002
LOCUS      BC030688
DEFINITION Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:26191
IMAGE:4822550, mRNA, complete cde.
ACCESSION   BC030688
VERSION     BC030688.1 GI:21265132
KEYWORDS
SOURCE      MGC.
ORGANISM    Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 3130)
Strausberg, R.
Direct Submission
Submitted (24-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshuyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>  
 contact: amadan@systemsbiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 33 Row: d Column: 5  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18450277.

FEATURES

source

1..3130

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:26191 IMAGE:4822550"

/cfeature\_type="Transcript"

/clone\_id="NIH\_MGC\_97"

/lab\_host="DH10B"

/note="vector: pBluescript"

372..2867

/codon\_start=1

/product="similar to dipeptidylpeptidase 8"

/protein\_id="AAH30688.1"

/db\_xref="GI:21265133"

/translation="MAAMETEQQLGVEIFETADCEENIESQDRPKLEPFYVERYSMSQ  
 LKRLAATRKHYGMAKAPHDPEFVRNDGDGSHRIYVLAMSGENREXTLPYSEI  
 PKTNRNAVLMISKRELLDFOALTDYGVNSREELRERRIGTGVIGASVDYHSGS  
 TFLFOASGTYHVYDGGPQGTQPLAPNLVETSCPNIRNDPKCPADPDAIATFHSN  
 DIWISNITVTRERKLTIVNHLEAMERDASAGATVLOEEFVRSYCMCPAETV  
 PSQKSLIRILENDESEVEIIVHTSPLELRADSRFRYPTGTANPVITKMEIM1  
 DAEGR1IDVIDKELIQPEILFEGVEIYIAGMPEKAYMSILDSQTLQVLIS  
 PELFIPEDVDMEORLIESVDSVTPLIYEETDIDIMNHDIFHYPQSHSEIER  
 IPASECTGRHLKYITSLKESKYRSGSLPAPSPKCP1KEEIAITSGEMVLGR  
 HGSNIQVDERKLVYFEGTQDSPLEHLLVYVSYNPSEVRLTRGRSHSCICQCHD  
 FFSKISNOKPHCVSLYKLSPPDDDFCTKREKWTATLDSAGLPDTPETSPES  
 TTGFTLLGSMYKPHLDLPGRKYPVLFIYGGPQVOLLNRRKGVYRFLNLAIGVY  
 VVVIDNRGSCHRGLKPEGAFKRYKVAIAGAIVTLMIVYDGTETERYMGHPQNEGYV  
 LGSVAMQAEKRPSEPNLLHGLFDENVHFATSIILSFLVRAGKPYDQIYQERH  
 SIRVEGHEHYELHLHYLOENIGSR1ALKVI"

BASE COUNT 955 a 626 c 717 g 832 t

ORIGIN

Alignment Scores:

Pred. NO.: 9,27e-233 Length: 3130

Score: 2447.50 Matches: 465

Percent Similarity: 90.47% Conservative: 0

Best Local Similarity: 90.47% Mismatch: 49

Query Match: 98.61% Indels: 1

Gaps: 1

US-10-070-464-5 (1-465) x BC030688 (1-3130)

Qy 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20

Db 1323 ACAGGTACAGCAATCTTAAGTCACTTTAAGATGTCAGAAATAAGATGAGCTGAA 1382

Qy 21 GlyArgIleIleAspValIleAspGluLeuIleGlnProPheGluIleLeuPheGlu 40

Db 1383 GGAAGATCATGATGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1442

Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60

Db 1443 GGAAGTTAATATATATGTCAGAGCTGATGATGATGATGATGATGATGATGATGAT 1502

Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80

Db 1503 CTACTAGATGCTCCACAGCTCCCTACAGATGATGATGATGATGATGATGATGATGAT 1562

Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100

Db 1563 CCACTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1622

QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTrrIleAsnIleHisAspIlePhe 120  
 DB 1623 AGCCCACTAATTATCTATGAGAAACACAGACATCGATTAATCCATGACATCTTT 1682  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAsnSerGluCysLys 140  
 DB 1683 CATGTTTTTCCCAAGTACAGAAAGAGAAATGAGTTTATTTTGGCTCTGAATGCATA 1742  
 QY 141 ThrGlyPheArgHisIleuTyrLysIleHisSerIleuLysGluSerLysTyrLysArg 160  
 DB 1743 ACAGGTTTCCGTCATTATTAACAATAATTAATTTTAAAGAAAGCAATTAACGA 1802  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleTyrGluGluIleAla 180  
 DB 1803 TCCAGTGGTGGCTGCTGCTCCAGATGATTTCAAGTCTCTTCAAAAGGAGATGCA 1862  
 QY 181 IleThrSerGlyGluThrGluValLeuGlyArgHisGlySerAsnIleGluValAspGlu 200  
 DB 1863 ATTACCACTGGTAAATGGGAAGTTCTTGCCGCGCATGATCTAATATCCAAAGTTGATGAA 1922  
 QY 201 ValArgArgLeuValTyrPheGluGluTyrThrLysAspSerProLeuGluHisIleuTyr 220  
 DB 1923 GTCAGAAAGCTGGTATATTTTGAAGGACCAAGACCTCCCTTTAGACATCACTGTAC 1982  
 QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 1983 GTAGTCAGTTACTTAATCTCGAGAGGTACCAAGCTGACTACCTGGCTACTACAT 2042  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsn 260  
 DB 2043 TCTTGCTGCATCAGTACGACGACTGACTCTTTATATAGTATAGTAAACGAAAGAT 2102  
 QY 261 ProHisCysValSerLeuTyrLysLysSerSerProGluAspAspProThrCysLysThr 280  
 DB 2103 CCACACTGGTGGTCCCTTTACCAAGCTATCAAGTCTGAAGATGACCACTTGCAAAACA 2162  
 QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 DB 2163 AAGGAATTTGGGCCCACTTTTGGATTCAGCAGTCTCTTCTGACTATCTCTCCCA 2222  
 QY 301 GlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 2223 GAATATTTCTCTTTGAAAGTACTACTGATTAACATTTGATGAGATGCTCAAGGCT 2282  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 340  
 DB 2283 CATGATCTACAGCTGGAAAGAAATATCTTAATGCTGTCTATATATGTTGTTCTCTCAG 2342  
 QY 340 ----- 340  
 DB 2343 GTGCAGTGGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTACCC 2402  
 QY 340 ----- 340  
 DB 2403 TCTCTAGTTATGTGGTTGATGATAGACAACAGGGGATCTGTACACGAGGGCTTAA 2462  
 QY 341 -----ValAlaIleAlaGlyAlaProValThrLeuTyr 351  
 DB 2463 TTTGAAGGGCTTTAAATATATAAATGTTGCTATTGCTGGGGCCCAAGTCACTGTGG 2522  
 QY 352 IlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGln 371  
 DB 2523 ATCTTATGATACAGATACAGACGAACTTATATGGGTCACTTACCAAGATGAAACAG 2582  
 QY 372 GlyTyrTyrLeuGlySerValAlaMetGlnIleGluLysPheProSerGluProAsnArg 391  
 DB 2583 GGCTATTACTAGATCTGTGGCCATGCAAGCAGAAAGTCTCCCTGAAACCAATTCGT 2642  
 QY 392 LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 411  
 DB 2643 TTACGTCTTACATAGGTTTCCGAGTGAAGATGTCATTGTGCAATACCAATATATTA 2702

QY 412 LeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGluIleTyrProGlnGluArg 431  
 DB 2703 CTGATTTTTTTAGTGAAGGCTGGAAAGCCATATGATTTACAGATCTATCTTCAGAGAGA 2762  
 QY 432 HisSerIleArgValProGluSerGlyGluHisTyrGluLeuHisIleuLeuHisIleu 451  
 DB 2763 CACAGCATTAAGAGTTCCTGATCGGAGAAACATTATGAACTGCATCTTTGCACTACTT 2822  
 QY 452 GlnGluAsnLeuGlySerArgIleAlaIleLeuLysValIle 465  
 DB 2823 CAGAAACCTTGATCAGATATGCTGCTCTAAAGTGATA 2864  
 RESULT 6  
 AX608731 4523 bp DNA linear PAT 17-FEB-2003  
 LOCUS AX608731  
 DEFINITION Sequence 8 from Patent WO0231134.  
 ACCESSION AX608731  
 VERSION AX608731.1 GI:28404301  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.  
 AUTHORS Novel serine protease genes related to dppiv  
 TITLE Patent: WO 0231134-A 8 18-APR-2002;  
 JOURNAL Ferring BV (NL)  
 FEATURES  
 Location/Qualifiers  
 source 1..4523  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 1384 a 828 c 940 g 1371 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1 93e-232 Length: 4523  
 Score: 2446.50 Matches: 463  
 Percent Similarity: 99.57% Conservative: 0  
 Best Local Similarity: 99.57% Mismatches: 1  
 Query Match: 98.57% Indels: 2  
 DB: 6 Gaps: 1  
 US-10-070-464-5 (1-465) x AX608731 (1-4523)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 1165 ACAGGTACAGCAAAATCTAAAGTCACTTTAAGATGTCAGAAATATATGATGATGCGGAA 1224  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 1225 GGAGGATCATGATGATCTCATAGTAAGAACTAATCAACCTTTAGATTCATTTGAA 1284  
 QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 1285 GGAATTTAAATATTTTCCAGAGCTGATGACTCCGAGGAGAAATATGCTGTGCATC 1344  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 1345 CTAAGTATGATCGCTCCAGACTGCGCTACAGATGATGATTCACCTGAATATTTATTC 1404  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 1405 CCAAGTAAAGATGATGTTTATGAAAGGACAGACTCATATGAGTCACTGCTGATTCGTG 1464  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTrrIleAsnIleHisAspIlePhe 120  
 DB 1465 AGCCCACTAATTATCTATAGAAACACACAGCATGATTAATATTCATGACATCTTT 1524  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAsnSerGluCysLys 140  
 DB 1525 CATGTTTTTCCCAAGTACAGAAAGGAGAAATGAGTTTATTTTGGCTCTGAATGCATA 1584



QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 DB 1585 ACAGGTTTCGCTCATTTATACAAATTCATCTATTTAAAGGAAGCAATTAACGA 1644  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluLysIleAla 180  
 DB 1645 TCACAGTGGTGGCTGCTGCTCCAAAGTATTCAGAGTCTCTCAAAAGGAGATGCA 1704  
 QY 181 IleThrSerGlyGluTTPGluValIleGluIYArgHisGlySerAsnIleGluValAspGlu 200  
 DB 1705 ATTACCAAGTGTGAATGGAAAGTTCTTGCCGCGCATGATCTAATATCCAAAGTTGATGA 1764  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220  
 DB 1765 GTCAGAAAGCTGGATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGTAC 1824  
 QY 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 1825 GTAGTGCATTCGTAATCTGAGAGAGTGAACAAGCTGACCTGCTGCTACTCAT 1884  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 1885 TCTTCTGCTCATGCTGACCTGACTCTTTTATAGTAAGTAAGTAACAGAAAT 1944  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 1945 CCACACTGTGTGCTCCCTTACAAAGCTATCAAGTCTGAAAGTGAACCACTGCAAAACA 2004  
 QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAlaPyrThrProPro 300  
 DB 2005 AAGGAATTTTGGGCGACCATTTTGGATTGACAGAGTCTCTTCTGCTACTCTCCCA 2064  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 2065 GAATTTTCTCTTTGAAATGACTACTGATTTACATTTAGATGCTCTACAGACT 2124  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 340  
 DB 2125 CATGATCTACAGCTGGAAGAAATATCTACTGCTGCTGCTATATGTGTGCTGCG--- 2180  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 2181 GTTGCTATTCCTGGGCGCCAGTCACTGTGATCTTCTATGATACAGATACACGGAA 2240  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 2241 CGTTATATGGGTACCTGACCAAGAAAGAACAGGGCTATTACTTAGAGTCTGTGGCATG 2300  
 QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 DB 2301 CAAGCAGAAAGTTCCCTCTGAAACCAATCGTTTACTGCTTACATGCTTCCCTGAT 2360  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyLys 420  
 DB 2361 GAAAGATGTCATTTGACATACCAAGTAATTACTGATTTTGAAGAGGCTGGAAG 2420  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 2421 CCATATGATTTTACAGATCTATCTCTCAAGAGACACAGCATTAAGAGTTCTGATCCGA 2480  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 DB 2481 GAACATTAATGAACTGATCTTTTGGACTACTTCAAGAAACCTTGATCATCATATTGCT 2540  
 QY 461 AlaLeuLysValIle 465  
 DB 2541 GCTCTAAAGTGATA 2555  
 RESULT 7  
 BD157001 2161 bp DNA linear PAT 17-JAN-2003  
 LOCUS BD157001  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.

. ACCESSION BD157001  
 VERSION BD157001.1 GI:27862759  
 KEYWORDS JP 2002191363-A/11844-  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 2161)  
 AUTHORS Oca,T., Isoqai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
 Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
 TITLE Primer for synthesizing full-length cDNA and use thereof  
 JOURNAL Patent: JP 2002191363-A 11844 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002191363-A/11844  
 PD 09-JUL-2002  
 'EP 28-JUL-2000 JP 2000280990  
 P1 TOSHIO OTA, TAKAO ISOQAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 P1 SAITO,  
 P1 JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 P1 KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH key  
 FEATURES  
 source location/Qualifiers  
 1..2161  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 FT CDS location/Qualifiers  
 (16)..(1911).  
 BASE COUNT 662 a 429 c 452 g 618 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2..2e-232 Length: 2161  
 Score: 2441.50 Matches: 464  
 Percent Similarity: 90.27% Conservative: 0  
 Best Local Similarity: 90.27% Mismatches: 1  
 Query Match: 98.37% Indels: 49  
 DB: 6 Gaps: 1  
 US-10-070-464-5 (1-465) x BD157001 (1-2161)  
 QY 1 ThrGlyThrAlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 370 ACAGGTACAGCAATCTTAAAGTCACTTTAAGATGTCAAGAAATTAATGATGATGCTGAA 429  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 430 GGAAGATCATAGATTCATAGTAAGTAAGAACTTAATTCMACTTTAGATTCATATTGAA 469  
 QY 44 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 490 GGAGTTGAATATATTCCTCAGAGCTGATGAGCTCTGAGGAAATATATGCTGTGCATC 549  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 550 CTACTAGATGCTCCCGACCTGCTTACAGATAGTGTATTCACCTGAAATTAATTAATC 609  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 610 CCACTGAAGATGATTTTATGGAAGGACAGACATCATTAAGTCATGCTGCTGATTTCTG 669  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 670 ACCGCCTAATTAATCATAGAGAAACACAGACATCTGATTAATATCATGACATCTTT 729  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 730 CATGTTTTCCTCCAAAGTACAGAGGAGAAATGAGTTATTTTTCCTCTGAAATGCAAA 789



QY 141 ThcIyPheargHieLeuYrLyseIlethSerIleLeuYsgIuseYrLyseYr 160  
 DB 790 ACAGGTTTCCTCATTTATACAAAATTCATCTATTTAAAGAAAGAAATATTAACGA 849  
 QY 161 SerSerGlycIleuProAlaProSerAaphelYsCySPoileYsgIuIleAla 180  
 DB 850 TCCAGTGGTGGCTGCCTGCCAAGATTCATTCAGTCTCTATCAAAAGAGATACGA 909  
 QY 181 IlethSerGlyGluTTPGluValLeuGlyYrghseGlySerAenIleGluValAepGlu 200  
 DB 910 ATTACCGATGGATGGAGAGGATCTTGCGCGCATGATCTAATATCCAGATTGATGATA 969  
 QY 201 ValArgArgLeuValIYrPheGluGlyThrIlysaSPSerProLeuGluHieIleuYr 220  
 DB 970 GTCAAGAGCTGATATTTTGAAGAGCCAAAGATCTCCCTTTAGAGCATCACCTGTAC 1029  
 QY 221 ValValSerYrValAenProGlyGluValThrArgLeuThrAaPaArgIYrSerHie 240  
 DB 1030 GTAGTCAGTACGTAATCTCGAGAGGTGACAAAGCTGACGACCGTACTACTCAT 1089  
 QY 241 SerYsCySileSerGlnHieCyAaSPhehIleSerIlyrSerAenGluYrAen 260  
 DB 1090 TCTTGCTGCATCAGTCAAGCTGTGACTCTTTATTAAGTATAGTATACAGAGAT 1149  
 QY 261 ProHieCySValSerLeuYrLyseYrLeuSerSerProGluAaPaProThCySLeYrThr 280  
 DB 1150 CCACACGTGTGTCTCTTTACAGCTATCAAGTCTCTGAAAGTGAACCACTTGCAAAACA 1209  
 QY 281 LysGluPheTTPAlaThrIleLeuAaSPSerAglYrProLeuProAaPyrThrProPro 300  
 DB 1210 AAGGAATTTGGGCGACCATTTTGATTCACAGAGTCTCTCTGACTATATCTCTCA 1269  
 QY 301 GluIlePheSerPheGlySerThrThcIyPheThrLeuYrGlyYmeIleYrLysePro 320  
 DB 1270 GAAATTTCTCTTTGAAAGTACACGATTTTACATTGATVAGAGTCTCTTACAGCT 1339  
 QY 321 HieAaPheGluInProGlyLyseYrProThrValLeuPheIleYrGlyGlyProGln 340  
 DB 1330 CATGATCTACAGCTGGAAGAAATATCTCTGTGCTCATATATGTGTGTCTCAG 1389  
 QY 340 ----- 340  
 DB 1390 GTGCGATTGGTAATAATCGGTTTAAAGAGTCAAGTATTTCCGCTGAATACCTTAAGC 1449  
 QY 340 ----- 340  
 DB 1450 TCTCTAGGTTATGTGTGTAGTAGACAAACAGGGATCTGTCAACGAGGCTTAAA 1509  
 QY 341 -----ValAlaIleAlaGlyAlaProValThrLeuTyr 351  
 DB 1510 TTTGAAGGCCCTTAATAATAAAATGTTCTATTGCTGGGCCCCAGTCACTCTGTGG 1569  
 QY 352 IlePheYrAaPThrGlyYrThcGluArgYrYrMetGlyHieProAaPGLnaenGluGln 371  
 DB 1570 ATCTTCAATATACAGATACACGAAACGTAATATGTGTGTCACTGTCAAGATGAAGCAG 1629  
 QY 372 GltYrYrIleuGlySerValAlaMetGlnAlaGluYrPheProSerGluProAaArg 391  
 DB 1630 GGGCATATCTAGGATCTGTGGCCATGACAGCAAAAGATCCCTCTGAAACCAAAATCGT 1669  
 QY 392 LeuLeuLeuLeuHieGlyPheLeuAaSPGLnaenValHiePheAlaHieSerIleu 411  
 DB 1690 TTAAGTCTCTTACATGTCTTCTCGATGAGAAATCTCATTTTGACATACAGATATTA 1749  
 QY 412 LeuSerPheLeuValAaGlaGlyYrProYrAaPLeuGlnIleYrProGlnGluArg 431  
 DB 1750 CTAGATTTTAAAGAGGCTGGAAGCAATATGATTTACAGATCTATCTCTCAAGAGAGA 1809  
 QY 432 HieSerIleArgValProGluSerGlyGluHieIleYrGluLeuHieLeuHieIleu 451  
 DB 1810 CACAGCATTAAGAGTTCTGAAATCGGAGAACATTTATGAATGCAATCTTTTGCACTACCTT 1869

QY 452 GlnGluAenLeuGlySerArgIleAlaLeuYsValIle 465  
 DB 1870 CAGAAAACCTTGATCACGATATTCCTCTTAAAGATGATA 1911  
 RESULT 8  
 AK027826  
 LOCUS  
 DEFINITION Homo sapiens cDNA FLJ14920 fis, clone PLACE1007416, weakly similar to DIPPTDYL PEPTIDASE IV (EC 3.4.14.5).  
 ACCESSION AK027826  
 KEYWORDS AK027826.1 GI:14042789  
 VERSIONS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Isega,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Magatsuma,M., Hosokari,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y., Nimomiya,K. and Iwayanagi,T.  
 TITLE NEBO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2161)  
 AUTHORS Isega,T. and Otsuki,T.  
 JOURNAL Direct Submission  
 COMMENT Submitted (10-MAY-2001) Takao Isega, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
 FEATURES  
 Source  
 Location/Qualifiers  
 1..2161  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="PLACE1007416"  
 /issue\_type="placenta"  
 /clone\_id="PLACE1"  
 /note="cloning vector: PMB18SFJ3"  
 16..1914  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB55395.1"  
 /db\_xref="GI:14042790"  
 /translation="MDPKLCPADBDWIAFIHSNDIWSINIVTREERLTYYHNELANM EEDARSAGVATFVLOEBFRDYGWMCPRKABETPSGGLILRIYEENDESVEIIVT SPLETRFARASFPYKCTGANTPVTTFPMSPSMDIDABERITIDIDKEILQPEILFECV EYLARAKMTPEGKYAMSTILIDRSQRLQIVLISPELPIPEDDMEMQRILIESPDSY TPULIBETDIDINIHIDIFHVEPOSHEIEIFLFAECKTGFEHLKYITSLKESY KRSSGKIPAPSDKCPKIEIATISGEWVILGRGSMIOVDEVARLVAFPGTQSDPE HHVYVVSYPVSEVTRITLDRYSVSHSCISQCHDFLSKYNOKRPHCVSYKXSSPED DPCKTKEFWATILDSAGPLPDVTPPEIFSFESTGTELLYGMILKPHDLPQKKXPVY LPYIGPOVOLVNNRPFQVYKVPFLNTLASIGYVVVVDNNGSCHRGKIKPGAPYKQV AINAGAPVTLMIFYDTGTERYKMHGPDONBEGYILGSAVMAEKRPSPNLLHGL DENVHFRVHTLSILSFLVRACKPYDLOLYPQERSIRVPSGEHYELHLHYLDENLGS RIALAKVI"  
 BASE COUNT 662 a 429 c 452 g 618 t  
 ORIGIN  
 Alignment Scores: \*  
 Pred. No.: 2,2e-232 Length: 2161  
 Score: 2441.50 Matches: 464  
 Percent Similarity: 90.27% Conservative: 0  
 Best Local Similarity: 90.27% Mismatches: 1

Query Match:	98.37%	Indels:	49
DB:	9	Gaps:	1

OY	1	ThrglyThrIlaAspProLysValThrPheLysMetSerGluIleMetIlaAspAlaGlu	20
Db	370	ACAGGTACAGCAAACTCTTAAAGTCACTTTTAAAGTGTCAAGAAATATATGATGATGATGCA	429
OY	21	GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu	40
Db	430	GGAAGGATCATAGATGATCATAGATAGAGAACAAATTCACCTTTTGAAATTCATTTGAA	489
OY	41	GlyValGluTyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIle	60
Db	490	GGAGTGAATATATTGGCCAGACTGGATGCACTCTGAGGGAATAATGCTGGTCCATC	549
OY	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle	80
Db	550	CTACTAGATCCCTCCCAAGACTCGCTCAAGATAGTGTTGATCTCCACCTGAATTTATTC	609
OY	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100
Db	610	CCAGTAGAAGATGATGTTATGGAAAGCAGAGACCTATGATGATGATGCTGATTCGTG	669
OY	101	ThrProLeuIleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePhe	120
Db	670	ACGCCACATTAATATGATATGAAGAAACAAGACATCTGGATTAATATCCAGACATCTTT	729
OY	121	HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys	140
Db	730	CATGTTTTCCTCCCAAGTCACAGAAAGAAATGTGTTATTTTGGCTCTGAATGCAAA	789
OY	141	ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg	160
Db	790	ACAGGTTTCCGTCATTTATACAAATTCATCATATTATTAAGAAAGCAAAATATTAACA	849
OY	161	SerSerGluGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla	180
Db	850	TCCAGTGGGGGCTCCTGCTCCAAAGTGAATTCAAAGTCTCATCAAGAAAGAGAGATACA	909
OY	181	IleThrSerGlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu	200
Db	910	ATTACCAAGTGGATGGAGAGTTCTTGCGCGGCATGATCAATATCCAGATTGATGAA	969
OY	201	ValAlaArgAlaGluValTyrPheGluGluLysHisAspSerProLeuGluHisIleLeuTyr	220
Db	970	GTCAGAAAGCTGGTATATTTGAAAGCAACCAAGACTCCCTTAAAGACATCACTGTAC	1029
OY	221	ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis	240
Db	1030	GTAGTCAGTACGTAAATCTGTGAGAGGAGCAAGCTGACGTGACGTAGCTACATCAT	1089
OY	241	SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn	260
Db	1090	TCTTGCTGCATCAGTCACACACTGTACTCTTTAATAAGTAAAGTAAAGAACAGAGAT	1149
OY	261	ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr	280
Db	1150	CCACACTGTGTCTCCCTTTACAAAGCTAACAGTCTGAAAGATGACCACACTGGAAAACA	1209
OY	281	LysGluPheTrpAlaIleThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro	300
Db	1210	AAGGAATTTTGGGCACCAATTTGGATTCAAGAGGCTCTCTCCGATCAATATCTTCCA	1269
OY	301	GluIlePheSerPheGluSerThrThrGlyPheTrpLeuTyrGlyMetLeuTyrLysPro	320
Db	1270	GAAATTTTCTCTTTGAAAGATCACTAGATTTAATCATTTGATGGATGCTCTCAAGACT	1329
OY	321	HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln	340
Db	1330	CATGATTCACAGCCGTGAAGAAAGAAATATCTACTGTGGCTGTCAATATAGAGGTCCTAG	1389

QY	340	-----	340
Db	1330	GTGCACTGGTGAATAATCGGTTAAAGAGTCAAGTATTTCCGTTGAATACCTTAGCC	1449
QY	340	-----	340
Db	1450	TCTCTAGGTATGTGGTTGTAGTATGACACACAGGGATCTCTGCACCGAGGCTTAAA	1509
QY	341	-----	341
Db	1510	TTTGAAGCGCCTTAAATATATAAATGATGCTGATATTCCTGGGGCCCAAGTCACTCTGTGG	1569
QY	352	IIePheTYraEPThrGlyTYrThGluARSTYrMeGlyHISProAspGlnAsnGluGln	371
Db	1570	ATCTTCTATATGATACAGGATACACGGAACGTTATATGGGTCAACCTGCACCAAGAAATGAACAG	1629
QY	372	GLYTYrTYrLeuGlySerValAlaIleGlnAlaGluLysPheProSerGluProAsnArg	391
Db	1630	GGCTATTACTTACCTAGGATCTGTGGCCATCCAGACAGAAAAGTTCCCTCTGAACCAATCGT	1689
QY	392	LeuLeuLeuLeuHISGlyPheLeuAspGluAsnValHISPheAlaHISThrSerIleLeu	411
Db	1680	TTACTGCTCTTACATGGTTCTTCTCGAGAGAAATGTCATTTTGACATACCAAGATATTTTA	1749
QY	412	LeuSerPheLeuValAlaArgAlaGlyLysProTYrAspLeuGlnIleTYrProGlnGluArg	431
Db	1750	CTGAGTTTATTTAGTACGAGGCTGGAACCAATATGATTTACAGATCTATCCCTCAGAGAGA	1809
QY	432	HISerIleArgValIProGlnSerGlyGluHISrTYrGluLeuHISLeuLeuHISrTYrLeu	451
Db	1810	CACAGCATTAAGATTCCTGGAATCGGAGAACATTATATACTGATCTTTTGGACTTACCTT	1869
QY	452	GlnGluAsnLeuGlySerArgIleAlaAlaLeuLysValIle	465
Db	1870	CAAGAAAACCTTGATACAGTATTTGCTGCTTAAAGATGATA	1911
RESULT 9			
LOCUS	AX354795	2649 bp	DNA linear PAT 06-FEB-2002
DEFINITION	Sequence 3 from Patent WO0179473.		
ACCESSION	AX354795		
VERSION	AX354795.1	GI:18619528	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE			
AUTHORS	Meyers, R.A. and Williamson, M.		
TITLE	21553, a human prolyl oligopeptidase family member and uses thereof		
JOURNAL	Patent: WO 0179473-A 3 25-Oct-2001;		
FEATURES	Millennium Pharmaceuticals, Inc. (US)		
source	Location/Qualifiers		
	1..2649		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	803 a 514 c 586 g 746 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2.51e-230	Length:	2649
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	92.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	6	Gaps:	1
US-10-070-464-5 (1-465) x AX354795 (1-2649)			
QY	1	ThrGlyThralaAsnProLysValIThrPheLysMeSerGluIleMetIleAspAlaGlu	20
Db	952	ACAGGTACAGCAAAATCTCTAAAGTCACTTTTAAGTGTGCAGAAATTAATGATTTGATGTGGAA	1011



/product="dipeptidyl peptidase IV-related protein-1"  
/protein\_id="AA017261.1"  
/db\_xref="GI:27549550"  
/translation="MAAMETEOGLVEIFETADCEENIESODRPKLEPFYVERYSMSO  
LKLILADRYKHYGMAMAKAPHDPMFVVRNPDGPHSRITVYLMSGRENPTLPSYRI  
PKTINRAVIMLSKPLLDLPQATLDYGMISREBELREKRGITVIAIADYHOSGSG  
TFLQAGSGIYHVADGPGQPTOPLFNLYVSCPIRNDPCPLCPDPMWIAFHSN  
DIWISNVTREERLTYVHNELEMEADASAVATLOEEDRSGYWCPEAET  
PSGKILRLIYEENDESEVEIHVTSPLMSTRADSRFRPRTGTANPKYKSEIM  
DAEGRIIDVIDKEILOPELIFEGVEYIAGMTPEGKYAMSILLDSOTRKLVLIS  
PELPIEDDMERORLIESVPSTPLIYEETDOWIMIHIFHVPQSHSEELER  
IFASECTGRFRLIKYTSIIKESYKRSSSGLRAPSPPKPIRELIATSGENVLGR  
HGSNIQDEVRRLVYFPGIKOSPLHHLIVYVSVNREAVRLDRGISHCISQHD  
FPISKYSNQNPICVSLYKLSPEDDPTCKEFWATIIISAGLPDYTPPELFSFES  
TTFGLYGMLYKPHDLOPKKYPVLIYGGPOVLVNNRFGKYPRLNTLASLGY  
VVVINDRGSCHRGKFEKAFYKMGQIEIDOVEGLQYLAHYDFIDLDRHGMGY  
GGVLSMALMGSDIPFRVALAGAPVTLMIYDTGYGTYWGHDPQNGYGLSVAMQ  
AKRPEPNNRLLIHGFLDENHFAHTSIIISFLVRAGKRYDLOIYQEHRSIRVES  
GEHYELHLHYLOENLGSRIALKYI"

BASE COUNT 804 a 517 c 588 g 747 t  
ORIGIN

Alignment Scores:

Pred. No.: 2,52e-230 Length: 2656  
Score: 2422.00 Matches: 465  
Percent Similarity: 82.30% Conservative: 0  
Best Local Similarity: 82.30% Mismatches: 0  
Query Match: 97.58% Indels: 100  
DB: Gaps: 1

US-10-070-464-5 (1-465) x AY172659 (1-2656)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 959 ACGAGTACACCAATCTTAAGTCATTTAAGATGCAGAAATTAAGATTGATGCGGAA 1018  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1019 GGAAGGATCATAGATGCATAGATTAAGAACTAATTAACCTTTGAGATTCTAATTGAA 1078  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTSPSerIle 60  
DB 1079 GGAATTGAAATTAATTCGCCAAGCTGATGATCTCTGAGGAAATATGCTTGCTCATC 1138  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1139 CTACTAGATCGCTCCAGACTCGCTGAGATAGTGTGATCTCACTGAAATTATTATC 1198  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1199 CCAGTAGAAATGATGTTATGAAAGGACAGACTCATGAGTCAAGTGCCTGATTCTGTG 1258  
QY 101 ThrProLeuIleIleTyrGluGluTyrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
DB 1259 ACCGCACTAATTATCTATGAAAGAAACAAGACTCTCGATTAATATCCATGACATCTTT 1318  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1319 CATGTTTTTCCCAAGTCAAGAAAGAAATGAGATTATTTTGGCTCTGAAATGCAAA 1378  
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
DB 1379 ACGAGTTTCGGCATTTATACAAATATACATCTAATTTTAAAGAAAGCAAAATTAACGA 1438  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
DB 1439 TCCAGTGGTGGGCTGCTGCCAAGTATTCAAGTGTCTCTATCAAAAGGAGATGCA 1498  
QY 181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1499 ATTAACAGTGGTGAATGGGAAGTCTTGCCGCGCATGATCTAATATCCAAAGTTGATGAA 1558  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluIleHisLeuTyr 220

DB 1559 GTGAGAGGCTGGATATTTTGAAGGACCAAAAGACTCCCTTAGAGCATCACCTGTAC 1618  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1619 GTAGTCAGTACGTAATCTCGAGAGGTACAAAGGCTGACTGACCTGGCTACTTCAT 1678  
QY 241 SerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
DB 1679 TCTTGCGCATCAGTACGACTGACTGCTTTTATAGTAAAGTATATCTAACCAAGAT 1728  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1739 CCACAGCTGTGCTCCCTTCAAGCATATCAAGTCCGMAATGACCCCACTTGCAAAACA 1798  
QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 1799 AAGGAATTTTGGGCGACCAATTTTGATTCAGCAGGCTCTTCTGACTATCTCTCCA 1858  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 1859 GAATTTTCTCTTTGAAAGTACTACTGATTTACATTTATGGAGTCTCTCAACACCT 1918  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
DB 1919 CATGATCTACAGCTCGAAAGAAATATCCATCTGTCTGTATATATGGTGTCTCAG 1978  
QY 340 ----- 340  
DB 1979 GTGCAATTGGTGAATATGATTTAAAGACGTAAGTATTTCCGCTGAAATACCCTAGCC 2038  
QY 340 ----- 340  
DB 2039 TCTCTAGTTATGTTGTTAGTATAGACAAACAGGGATCTGTACAGAGGCTTAAA 2098  
QY 340 ----- 340  
DB 2099 TTTGAAGCGCTTTAAATATTAATAATGGTCAATAGAAATGACGATCAGTGGAAACA 2158  
QY 340 ----- 340  
DB 2159 CTCGAATATCTACTCTTGATATGATTTCAATTGACTAGATGCTGTGGCATCCAGCC 2218  
QY 340 ----- 340  
DB 2219 TGTCTCATGAGGATACCTCTCCGTGATGACATTAATGACAGAGTCAAGATCTTCAGG 2278  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
DB 2279 GTTGCTAATGCTGGGCGCCAGACTCTGTGATCTTCTATGATACAGATACACGGAA 2338  
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
DB 2339 CGTTATATGGGTACCTCGACCAAGATGAACAGAGGCTATTACTTAGATCTGTGGCCATG 2398  
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
DB 2399 CAAGCAGAAAAGTTCCCTCTGAAACCAATCGTTTCTGCTCTTACATGCTTTCCTGAT 2458  
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 2459 GAGAAATGTCATTTTCCATACATACAGATATATTAAGTATTTTATGAGAGGCTGGAAG 2518  
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
DB 2519 CCAATGATTTACAGATCTATCTCTCAGAGAGACACAGCTTAAGATTTCTGATTCGGA 2578  
QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLysGlnGluAsnLeuGlySerArgIleAla 460  
DB 2579 GAACATTATGAAGTCACTCTTTTGCACTACCTTCAAGAAAACCTTGATCACTATTTGCT 2638  
QY 463 AlaLeuLysValIle 465



Db 2579 GAACATTAATGACATCTTTTGACCTTCAAGAAACCTTGATCACTATTGCT 2638

Qy 461 AIAleuysValIle 465  
 |||||  
 Db 2639 GCTCTAAAGTGATA 2653

RESULT 12  
 AX342633 3106 bp DNA linear PAT 12-JAN-2002  
 LOCUS Sequence 30 from Patent WO0198468.  
 DEFINITION AX342633  
 ACCESSION AX342633  
 VERSION AX342633.1 GI:18152030  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 1 Yue H., Elliott V.S., Gandhi A.R., Lal P., Au-Young J.,  
 Tribouley C.M., Delegeane A.M., Baughn M.R., Nguyen D.B., Lee E.A.,  
 Hatalla A., Khan F.A., Walla N.K., Iao M.G., Lu D.A., Patterson C.,  
 Tang Y.T., Walsh R.T., Azimzal Y., Ramkumar J., Xu Y. and Reddy R.  
 Patent: WO 0198468-A 30 27-DEC-2001;  
 Incyte Genomics, Inc. (US)

FEATURES  
 source 1..3106  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /note="Incyte ID No: 7160544CB1"

BASE COUNT 928 a 633 c 704 g 841 t

ORIGIN

Alignment Scores:  
 Prod. No.: 3.13e-230 Length: 3106  
 Score: 2422.00 Matches: 465  
 Percent Similarity: 82.30% Conservative: 0  
 Best Local Similarity: 82.30% Mismatches: 0  
 Query Match: 97.58% Indels: 100  
 DB: 6 Gaps: 1

US-10-070-464-5 (1-465) x AX342633 (1-3106)

Qy 1 ThrglythrAlaaspProValThrpheylMetSerGluIleMetIleaspAlaGlu 20  
 |||||  
 Db 1154 ACAGGTACAGCAATCTTAAGTCACTTTAAGATGTCAGAAATGATTGATGTCGAA 1213

Qy 21 GlyArgIleIleaspValIleaspIleGluIleGlnProPheGluIleLeuPheGlu 40  
 |||||  
 Db 1214 GGAAGGATCATAGTGCATGATAGGAACTAATTCACCTTTGAGATTCTATTGAA 1273

Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 |||||  
 Db 1274 GGAATTAAATATATGTCAGAGCTGATGACCTCTGAGGAAATATGCTTGCTCATC 1333

Qy 61 LeuLeuaspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 |||||  
 Db 1334 CTACTAATAGCTCTCCAGACCTCGCTACAGATAGTGTGATCTCACTGAATTAATTATC 1393

Qy 81 ProValGluaspArgValMetGluArgGlnArgLeuIleGluSerValProaspSerVal 100  
 |||||  
 Db 1394 CCAAGTAGAATGATGTTATGAGAAAGCAGAGACTCATTTGAGTCAGTCCGATTTCTGTG 1453

Qy 101 ThrProLeuIleIleTyrGluGluThrThraspIleTyrPheIleLeuIleHisaspIlePhe 120  
 |||||  
 Db 1454 ACCGCACTAATATCTATGAAAGAAACAACAGACATCGAATAATATCCATGACATCTTT 1513

Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysGly 140  
 |||||  
 Db 1514 CATGTTTTTCCCAAGTCAAGAGAGAAATTAAGTTATTTTGGCTCTGAATGCAAA 1573

Qy 141 ThgGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 |||||

Db 1574 ACAGTTTCCGTCACTTATACAAATTAATCATCTATTTTAAAGAAAGCAAAATATAACGA 1633

Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 |||||  
 Db 1634 TCCAGTGTGGGCTGCTGCTCCAAAGTATTTCAAGTGTCTTATCAAAAGAGATAGCA 1693

Qy 181 IleThrSerGlyGluTyrPheGluValLeuGlyValArgHisGlySerAsnIleGlnValaspGlu 200  
 |||||  
 Db 1694 ATTACCAAGTGTGAATGCGAAGATTCTGGCCGCGATGATTAATATCAAGTTGATGAA 1753

Qy 201 ValArgArgLeuValTyrPheGluGlyThrLysaspSerProLeuGluHisIleLeuTyr 220  
 |||||  
 Db 1754 GTCAAGAGGCTGATATTTTGAAGGCAACAAAGACTCCCTTTAGAGCATCACCCTTGAC 1813

Qy 221 ValValSerTyrValaaspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 |||||  
 Db 1814 GTACTCAGTTACGTAATATCCGAGAGGTGACAGGCTGATGACCGTGGTACTACAT 1873

Qy 241 SerTyrCysIleSerGlnHisCysaspPheIleSerLysTyrSerAsnGlnLysAsn 260  
 |||||  
 Db 1874 TCTTGCTGCATCACTCAGCACTGACCTGACTCTTTATTAAGTAAATAGTAACCAAGAAAT 1933

Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluaspProThrCysLysThr 280  
 |||||  
 Db 1934 CCACACTGTGTGCTCTTTACAGCTATCAAGTCTTAAGATGACCAACTTGCAAAACA 1993

Qy 281 LysGluPheThrPalatThrIleLeuaspSerAlaGlyProLeuProaspTyrThrProPro 300  
 |||||  
 Db 1994 AAGGAATTTTGGGCAACATTTTGATTCAGCAGGTCTCTCTGACTAATACCTTCCA 2053

Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 |||||  
 Db 2054 GAATTTTCTTTTGAAGTAACTACGATGATTCATTTGATGAGATGCTCTCAACACCT 2113

Qy 321 HisaspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 |||||  
 Db 2114 CATGATTTACAGCTCTGAAAGAAATATCTCACTGTGCTGTTCAATATAGGTGCTCAG 2173

Qy 340 ----- 340

Db 2174 GTGCAGTTGGTAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTAGCC 2233

Qy 340 ----- 340

Db 2234 TCTCTAGATTATGTGTTGTAGTATGATAGACACAGGAGATCTGTCAACGAGGCTTAAA 2293

Qy 340 ----- 340

Db 2294 TTTGAAAGCGCCTTTAAATATTAATGAGTCMAATAGAAATTGACGATCAGTGAAGGA 2353

Qy 340 ----- 340

Db 2354 CTCCAATATCTAGCTTCTCGATATGATTTCAATGACTTAGATCGTGGGATCCAGGC 2413

Qy 340 ----- 340

Db 2414 TGGTCTATGAGAGATACCTCTCCCTGANGCATTTAATGACAGGTCAGATATCTTGAG 2473

Qy 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 |||||  
 Db 2474 GTTGCTATTGCTGGGGCCAGTCACTCTGTGATCTTCTATATACAGATACAGGAA 2533

Qy 361 ArgTyrMetGlyHisProaspGlnaaspGluGlyTyrTyrLeuGlySerValAlaMet 380  
 |||||  
 Db 2534 CGTTATATGGGTCACTTCAGCAAGATGAAACAGGGCTAATTAAGGATCTGGGGCATG 2593

Qy 381 GlnAlaGluLysPheProSerGluProaspArgLeuLeuLeuHisGlyPheLeuasp 400  
 |||||  
 Db 2594 CAAGCAGAAAGTTTCCCTCTGAACCAAAATGTTTACTGCTCTTAACATGGTCTTCGAT 2653

Qy 401 GluaspValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 |||||  
 Db 2654 GAGAAATGTCATTTTGCACATACCAATATTAATTAAGTTTATGAGGGCTGGAAG 2713



QY	421	ProTyraapleagunilileTyProgingunurghisseriesileagValProguSercly	440
Db	2714	CCATTGATTTTACAGATCTTACTTCCAGAGAGACACACATTAAGATTCCTGGAATCGGGA	2773
QY	441	GlunhistryguleuHsileuLeuHsistryleugingunlaueLeuglySerArg1lea1a	460
Db	2774	GACACTATGAACTGCATCTTTGGCACTACCTTCAGAAAACCTTGATCAGTATTCT	2833
QY	461	AlaleuLyysValille	465
Db	2834	GCTCTAAAGTGATA	2848
RESULT 13			
AF221634			
LOCUS	AF221634	3127 bp	mRNA linear PRI 05-NOV-2000
DEFINITION	Homo sapiens dipeptidyl peptidase 8 (DPP8)	mRNA, complete cds.	
ACCESSION	AF221634		
VERSION	AF221634.1	GI:11095187	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 3127) Abbott,C.A., Yu,D.M., Woollett,E., Sutherland,G.R., McCaughan,G.W.		
TITLE	Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homologue, DPP8		
JOURNAL	Eur. J. Biochem. 267 (20), 6140-6150 (2000)		
MEDLINE	20467194		
PUBMED	11012666		
REFERENCE	2 (bases 1 to 3127) Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.		
AUTHORS	Direct Submission		
TITLE	Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver Centre, Centenary Institute of Cell Biology and Cancer Medicine, Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia		
JOURNAL	Location/Qualifiers		
FEATURES			
source	1..3127		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="15"		
	/map="15q22"		
	1..3127		
gene	/gene="DPP8"		
CDS	214..2862		
	/gene="DPP8"		
	/note="serine protease; member of prolyl oligopeptidase 89b family belonging to enzyme clan SC; alternatively spliced forms of this mRNA are found in AF221635, AF221636 and AF221637"		
	AF221637"		
	/codon_start=1		
	/product="dipeptidyl peptidase 8"		
	/protein_id="AAG29766.1"		
	/db_xref="GI:11095188"		
	/translation="MAAAAEYEQGVLEFPADEENIESODRPLEPFYVERYSNQ LKLLADLRKTHGYMAAPIDFEMFVRKNDPDSORITYYLAASGRENNTLYSEI PKTINRAVALMLSWKPLDLEQATLDYGMSREBELRERKIGTVAIASDYHOGSG TPLFQASGIGYHVDGDPGQFTQOPLRPNLVETSCPNIRMDPKCPADPMIAFIHSDIATSNIVTREERLTGYVHNELAMEDASAGATVLOEERFVSGYMMCPKAETT PSQGIKRLILEEYNDSEVELIHNTSMLETRPADSRVYRKTGPNKYMPEKSEIMT DABERRITDVIDKELIQPELIFBEVETIARAGWTPBEKTKAMSTILDSQRLLQIVLS PELTIPVEDDWMERKRLSEVDSVTPLEETTDIWINIHDFVFPQSHSEELF IIPAECHTFRHLKITYTSLKESYRKSSGGLPAPSDFKPIKEELIITSGEVELGR HGSNIQVDEVARLRYFEGTKDPSPLEHLIYVVSVPGEVTRLIDRGVSGSCISQHD FIFIKYENQKNPKHCVSLYKLSPEDDPTCKTKEFWATILDSAGLPDYTPPELIFSFS TTGTLTGMLTGKPHLDLOPKKYKPTVLFIYGGQVQVLNNRPKGVKPYRLATLSIGV VVIVDNGSGCHRGKLFKEGAFKYKMGQIEIDDOVSGIOLASRYDFIDLDVGIHGWY GGYSLIALMQRSDIPRYVALNAGATVTMTIPFDGYTTERYVGHGPNQEGYLLSVAMQ AEKPSPPNRLLDLHGFUDENVHFARHSTILSLFVIRGKPYDLOIYPOERHSIRVPS GEHHEHLLHYLLOENIGSRIALKVI"		

BASE COUNT	943 a	637 c	706 g	841 t
ORIGIN				
Alignment Scores:				
Pred. No.:	3,16e-230	Length:	3127	
Score:	2422.00	Matches:	465	
Percent Similarity:	82.30%	Conservative:	0	
Best Local Similarity:	82.30%	Mismatches:	0	
Query Match:	97.58%	Indels:	100	
DB:	9	Gaps:	1	
US-10-070-464-5 (1-465) x AF221634 (1-3127)				
QY	1	ThrgLYThrrAlaAspProLyValThrPhelYsmtSerGluIlleMetIlleAspAlaGlu	20	
DB	1165	ACAGGTACAGCAAAATCCTTAAGTCACTTTTAAGATGTCAAGAAATATATGATTATGCTGAA	122	
QY	21	GIlyArgIlleIleAspValIlleAspLySglubLeuIleGlnProPhelGluIlleuPhelGlu	40	
DB	1225	GGAAAGATCATVAGATGTCATGATGAATGAACCTAAATCAACCTTTTGAGATTCTATTGGA	128	
QY	41	GIlyAlaGIuTYrIlleAlaArgAlaGlyTYrPhrProGIuGIlySlyTYrAlaTYrSerIlle	60	
DB	1285	GGAGTGAATATATTTGCCAGACGTGAAGAGACTCCTGAGGGGAAATATGCTTGCTGCATC	134	
QY	61	LeuIleuAspAPrSergInThrArgLeuGlnIlleValIlleSerProGIubepheIlle	80	
DB	1345	CTACTAGATCCGCTCCCAACATCGCTTACAGATAGTGTGATCTCACCTGAATATTATTC	140	
QY	81	ProValGIuAspAspValMetGIuArgGlnArgLeuIlleGIuSerValProAspSerVal	100	
DB	1405	CCAGTAGAAGATGATGTTATGAAAGGACAGAGACTCATGTGATCGTGCCTGATTCCTGG	146	
QY	101	ThrProIleuIlleIleTYrGIuGIuThrThrAspIleTPIleAsnIlleHisAspIlePhe	120	
DB	1465	ACGCCACTAATATCTATGAAGAAACAACAAGCATCTGAATAAATATCCATGACATCTTT	152	
QY	121	HisValPheProGlnSerHisGIuGIuGluIleGIuPheIllePheAlaSerGIuSlyS	140	
DB	1525	CATGTTTTTCCCAAGACGACGAAGAGAAATTGATTTATTTTGCTCTGAATGCAAA	158	
QY	141	ThrgIyPheAlaGhIleuTYrLySIlleThrSerIlleuLySgluSerLySlyTYrLySArg	160	
DB	1565	ACAGTTTCCGCTCATTTATACAAATTAACATCTATTTAAAGAAACAAATATAACGA	164	
QY	161	SerSerGIyGIlyLeuProAlaProSerAspPheLySAspProIlleLySgluGluIlleAla	180	
DB	1645	TCCAGTGTGGAGCTGCCCTGCCACAGATATTTCAAGTCTCTATCAAAAGAGAGATAGCA	170	
QY	181	IlleThrSerGIyGIuTYrPIuValleuGIyAArgHisGIySerAsnIlleGlnValAspGIu	200	
DB	1705	ATTACACAGATGTGAATGGGAAGTTCTTGCGCGGATGTGATCTAATATCCAAAGTTGATGA	176	
QY	201	ValArgArgLeuValTYrPheGIuGIyThrLySAspSerProLeuGlnHisIleuTYr	220	
DB	1765	GTCGAAAGGCTGATATTTGAAAGGCAACCAAGACTCCCTTTAAGACATCACCTGTAC	182	
QY	221	ValValSerTYrValAsnProGIyGIuValThrArgLeuTYrAspArgGIyTYrSerHis	240	
DB	1825	GTAAGTCAAGTACGTAATCCCTGGAGAGGTGACAAAGCTGACGACCGTGGCTACTACAT	188	
QY	241	SerCySAspIlleSergInHisCyAspPheheIlleSerLySlySerAsnGlnLySAsn	260	
DB	1885	TCTTGCTGCATCAGCTGACGACCTGTGACTCTTTTAAAGTAAAGTAAACCAAGAAAT	194	
QY	261	ProHisCyAspValSerLeuTYrLySLeuSerSerProGIuAspAspProThrCySlySphr	280	
DB	1945	CCAACAGTGTGTCTCTTACAGCTATCAAGTCTGAAAGTGACCAACTGGCAAAACA	200	
QY	281	LySgluPheTYrAlaThrIlleLeuAspSerAlaGIyProLeuProAspTYrThrProPro	300	
DB	2005	AAGGAATTTTGGGCAACCATTTTGATTCAGCAAGTCTCTTCTGACTAATATCTCTCA	206	



Qy	301	Glul1ePheSerPheGluSerThrThrGlyPheThrLeuTyrglyMetLeuTyrlYLeuPro	320
Db	2065	GAATTTCTCTTGAAGAGTACTGATGATTTACATTTGATGGATGCTTACAAAGCTT	2124
Qy	321	HisAlaPheLeuGlnProGlyLysLysLeuTyrlProthrValLeuPheIleTyrglyGlyProGln	340
Db	2125	CATGATCTACAGCTGGAAGAAATATCTCATCTGTGCTGATATATGAGTCTCAG	2184
Qy	340	-----	340
Db	2185	GTGAGTTGGGTAATAATCGTTTAAAGAGTCAAGATTTCCGCTTGAATACCTTAGCC	2244
Qy	340	-----	340
Db	2245	TCTCTAGGTTATGTGTTGATGATAGACAAACAGGAGATCTGTCAACGAGGCTTAAA	2304
Qy	340	-----	340
Db	2305	TTTGAAGCGCCCTTAAATATAATAGGTCATAATAGAAATTGACATCAGGTGAAAGA	2364
Qy	340	-----	340
Db	2365	CTCCAATATCTAGCTTCCGATATGATTTCTTGACTTAAATCGTGTGGGATCCAGGC	2424
Qy	340	-----	340
Db	2425	TGCTCTATGAGAGATCTCTCCCTGATGGCATTAATGACAGAGTCAATATCTTACG	2484
Qy	341	ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrlAspThrGlyTyrlThrGlu	360
Db	2485	GTTCATATCTGCGGGCCCAAGTCACTGTGATCTTATGATACAGATACAGAA	2544
Qy	361	ArgTyrlMetGlyHisProAspGlnAsnGlnGlnGlyTyrlTyrlLeuGlySerValAlaMet	380
Db	2545	CGTTATATGGGTACCTCGAACCAAGATGAACAGGCGATTAATCCTAAGATCTGTGCCATG	2604
Qy	381	GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp	400
Db	2605	CAAGCGAAGAAAGTCCCTCGAACCAATCGTTACTGCTTACATGATGTTCTCGAT	2664
Qy	401	GluAsnValHisPheAlaHisSerThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys	420
Db	2665	GAGATATCTCAATTTTGCATACATCCAGATATATTACTGATGTTTTTAAAGAGGCTGGAAG	2724
Qy	421	ProTyrlAspLeuGlnIleTyrlProGlnGluArgHisSerIleArgValProGluSerGly	440
Db	2725	CCATATGATTTTACAGATCTATCCCTCAGAGAGACACAGCATTAAGATTCGGAATCGGA	2784
Qy	441	GluHisTyrlGluLeuHisLeuLeuHisTyrlLeuGlnGlnAsnLeuGlySerArgIleAla	460
Db	2785	GAAATATATGAACGATCTTTTGGACTACCTTCAGAAAAAACCCTTGATCAGCTATGCT	2844
Qy	461	AlaLeuLysValIle 465	
Db	2845	GCTCTAATAAGTGATA 2859	
RESULT 14			
AX354793	3143 bp	DNA	linear PAT 06-FEB-2002
LOCUS	Sequence 1 from Patent WO0179473.		
DEFINITION	AX354793		
ACCESSION	AX354793.1	GI:18619526	
VERSION			
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1		
AUTHORS	Meyers, R.A. and Williamson, M.		
TITLE	1		
JOURNAL	Patent: WO 0179473-A 1 25-OCT-2001;		
	Millennium Pharmaceuticals, Inc. (US)		
FEATURES	Location/Qualifiers		

SOURCE		1. 3143	/organism="Homo sapiens"
			/mol_type="genomic DNA"
			/db_xref="taxon:9606"
CDS		229..2877	/note="unnamed protein product"
			/codon_start=1
			/protein_id="CAD22990.1"
			/db_xref="GI:18619527"
			/translation="MAAAMETEQLGVEIFETADCEENIESODRKLPEPFYERSMSQ
			LKTLADTRKYHGMMAKAPHDPMFVBNRPPDPSHRIRYVLMASGSRNRLTYSGSI
			PTLNRAVLMLSKRLDLFQATLDYGMNRSREBLREKRIGTCTVIAIYDHQSG
			TLFLQAGSGIYHVKGDPQFTQPPNPVNSCPNIRNDPKLCPADPMIATHSN
			DIWISNIVTEERRLLTYVNHMLNMEBDAASGAFVLOEEDRSGVMSCPRAET
			PSGKILRIIYEENDESEVEIIVTSPMLTEBRADSERYPKGTANPKYFKMSIEM
			DARGRIIDIVDKELIOPFEILFEGVEYIAAGMTPEGVAMSIILDSORTLIVLS
			PELFPVEDVMEROLIESPDSVPLIYEETTDWINIHIFHVPQSHIEIEF
			IFASBECKTFRHLKRTSLIKESKYSRSGGLAPDPKPIREIAITSGEMVLR
			HGSNIQVDEVRRLVFEGETKDSPLRHHLLVYVSVNPEVRLTDRGYSHCISQHD
			FFISKYSNQKPHCVSLYKLSPEDDPTCKEFWAITLDSAGPLDPYTPPELFSFS
			TTGFTLYGMLYKPHDLOPKKYPVLFIYGGPQVQVNNRFKGVYRLNTLTLGIV
			VVVIDNRGSHRGLKFGGAFKYMGGQIEIDVGGYQYLASRYDFIDLDRVGHGMY
			GGYLSMALMORSDIRVATAGAPLYLMIFPYDGYRBYVGHGPDNONGCYLVAMQ
			AEKPSPEPNRLILHGLFDENVFPAHTSILSLVRAKPYDLQITPOENHSIRVPS
			GEHYELHLHYLDENLGSRIALKLV"
BASE COUNT		943 a 644 c 712 g 844 t	
ORIGIN			
Alignment Scores:			
Pred. No.:		3,18e-230	Length: 3143
Score:		2422.00	Matches: 465
Percent Similarity:		82.30%	Conservative: 0
Best Local Similarity:		82.30%	Mismatches: 0
Query Match:		97.58%	Indels: 100
DB:		6	Gaps: 1
US-10-070-464-5 (1-465) x AX354793 (1-3143)			
Qy	1	ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu	20
Db	1180	ACAGGTACAGCAATCTTAAGTACACTTTTAAGATGCAGAAATTAAGATTGAGCTGAA	1239
Qy	21	GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu	40
Db	1240	GGAAGGATCATATGATGCATAGATTAAGAACTAATCACTTTTGAGATTCATTGAA	1299
Qy	41	GlyValGluLysTrpIleAlaArgAlaGlyTrpThrProGlnLysTyrlAspIleAspIle	60
Db	1300	GAGTTGAATATATTCGCCAGAGCTGAGTGGACTTCGAGGAAATATATGCTGCTATC	1359
Qy	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle	80
Db	1360	CTACTAGATCGCTCCAGACTCCCTGCAGATAGTGTGATCTCACCCTGAATTTATTC	1419
Qy	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100
Db	1420	CCAGTAGAAGATGATGTTATGAGAAAGCAGACCTCATGACAGTCTGATTCGTG	1479
Qy	101	ThrProLeuIleIleTyrlGluGluTrpThrAspAlaTrpIleLeuIleHisAspIlePhe	120
Db	1480	ACGCCACTAATTTATATGAGAAACACAGACATCTGGAATATATCCATGACATCTTT	1539
Qy	121	HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys	140
Db	1540	CATGTTTTTCCCAAGATCAGAAAGGAAATGAGTTATTTTTCCTCTGAATGAA	1599
Qy	141	ThrGlyPheArgHisLeuTyrlLysIleThrSerIleLeuLysGluSerLysTyrlArg	160
Db	1600	ACAGGTTTCCGTCATTATACAAATTAATCAATTTTAAAGGAAAGCAATATTAACGA	1659
Qy	161	SerSerGlyGlyLeuProLysProSerAspPheLysCysProIleLysGluGluIleAla	180
Db	1660	TTCAGTGGTGGGCTGCTGCCAAGTATTTCAAGTGTCTTCAAGAGGAGATAGCA	1719

QY 181 ILeThSerGlyGluTTPGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 1720 ATTAACGAGTGGTGAATGGGAAGTCTTGCCGGCATGGATTAATATCCAAAGTTGAGAA 1779  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrIysAspSerProLeuGluHisIleLeuTyr 220  
 DB 1780 GTCAAGAGCGGTGATATTTTGAAGGACCAAGACATCCCTTTAGAGCATCACTGTAC 1839  
 QY 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 1840 GTAGTCATGTAAGTAATCCGAGAGGTGACAAAGCGTAGCTGACCGGTGCTACTCAT 1899  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsn 260  
 DB 1900 TCTTGGCATCATCACTGACACTGTGACTCTTTTAAGTAAGTATAGTAACCAAGAAAT 1959  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLysThr 280  
 DB 1960 CCAACACTGTGTCTCTTACAAAGCTATCAAGTCTCAAGATGACCAACTTGCAAAACA 2019  
 QY 281 LysGluThrPheTyrPheIleLeuAspSerIleGlyProLeuProAspTyrThrProPro 300  
 DB 2020 AAGGAATTTTGGGCGACCATTTTGATTCAGACAGTCTCTTCTGACTATATCTCTCCA 2079  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 2080 GAAATTTTCTCTTTGAAAGTACTAGTGAATTTACATTTGATGGAGTCTCTACAAAGCT 2139  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 2140 CATGATCTACAGCTGGAAAGAAATATCTACTGTGCTGCTCATATATGGTGGTCTCAG 2199  
 QY 340 ----- 340  
 DB 2200 GTGCGATTGTGAATAATCGGTTTAAAGAGTCAATATTTCGCTGAATACCTTACC 2259  
 QY 340 ----- 340  
 DB 2260 TCTCTAGTATATGTGTGTAGTATAGACAAAGGGATCTCTACCAAGGGCTTAAA 2319  
 QY 340 ----- 340  
 DB 2320 TTTCGAAGCGCCTTAAATATATAAATGGGTCAATAATGAAATTGACGATCAGTGAAGA 2379  
 QY 340 ----- 340  
 DB 2380 CTCCAATATCTAGCTTCTGATATGATTTCAATTGACTTAATCGTGGGCAATCCAGC 2439  
 QY 340 ----- 340  
 DB 2440 TGGTCTATGAGGATACCTCTCCCTGATGGCATTATGACAGAGTCAAGATATCTTACG 2499  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 2500 GTTGCTATTTCTGGGGGCCAGTCACTCTGGATCTTTATATGATACAGATACACGGA 2559  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 2560 CGTTATATGGGTCACTGACCAAGAAATGAACAGGGCTATTATCTTAAGATCTGTGGCCATG 2619  
 QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuHisGlyPheLeuAsp 400  
 DB 2620 CAAGCAGAAAAGTTCCCTCTGAAACAAATCGTTACTGCTTACATGGTTTCTTGAT 2679  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 2680 GAAATATTCATTTTGCACATACACAGTATATTAAGTTTGAAGAGGGCTGGAAG 2739  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 2740 CCAATATGATTTACGATCTATCTCTCAGAGAGACACAGCATTAAGAGTTCTCGAATTCGGA 2799

QY 441 GluHisTyrGluLeuHisIleLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 DB 2800 GAACATATATGACGATCTTTTGCATCTTACAAAGAAACCTTGATCAAGTATTCCT 2859  
 QY 461 AlaLeuLysValIle 465  
 DB 2860 GCTCTAAAGTGATA 2874  
 RESULT 15  
 BC040203  
 LOCUS  
 DEFINITION  
 Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:42570  
 IMAGE:4824813, mRNA, complete cds.  
 BC040203  
 VERSION  
 BC040203.1 GI:26007916  
 KEYWORDS  
 MGC  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 4535)  
 Straubeberg, R.  
 Direct Submission  
 Submitted (27-NOV-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovics, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shihaki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Guinarte, F.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Wuzny, D.M., Nanaavati,  
 A.N., Gibbs, R.A.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>  
 Series: IRAK Place: 71 Row: K Column: 23.  
 Location/Qualifiers  
 1. 4535  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:42570 IMAGE:4824813"  
 /tissue\_type="Testis"  
 /clone\_id="NH MGC\_97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 1629..4277  
 /codon\_start=1  
 /product="similar to dipeptidylpeptidase 8"  
 /protein\_id="AAH40203.1"  
 /db\_xref="GI:26007917"  
 /db\_xref="locusid:54878"  
 /translation="MAAAETBOLGVEIFETADCENISQDRPKLEPPYVERYSNQ  
 LKTLADTRKTYGWMAPKPHDFVAKRNDPDSHSDRIYLLAMSGENRNTLFSBI  
 PKTINRAVALMSKRLDLFOATLDYGVYSREEDLREKRIGTVGASVDYHSGG  
 TPLFOAGSGIYHVDCGPOGTQPLRPMLVETSCNIRMDPKLCPADPMIYFIHGN  
 DIWISNIVTREERRLTYVHNELANMEDRASAVATFVLOEEDRSGYWCRAET  
 PSGGRTILILENDESEVEITHTVSPMLSTRADSFRYKTKTAPKTPKSEIMI  
 DAGRTIYIDKELIOPREILPEGCYTIARAGTTPKAKMSILRSQTRILVLIS  
 PELFIVEDVWERQRLIESVPSVPLIIEETTDIWINIHIFVFPQSHSEIEF  
 IFASEKTEGRHLVYKTSILKESKSGGLPAPDPKPIKEIALISGSEVIGR  
 HGSNIQVDEVRRLVPEGTKDSLLEHLHLYVSVYNGEVRLTRDRGYSKSCQSD  
 FPKSYNQKPHCVSLYKLSPEDDPTCKTEFMTILDSAGPLDPYMPPELFSFS

TTGFTLYGMLYKPHDLPGRKYPVTLVLYGGPOVLVNNRKKYKPRRLNTLASLGV  
 VVVIDNRGSSCHRGKFEQAFRYKKGQLEIDVOVGLDYLSRYPFLDLDRGHGWXY  
 GGYLSMALMQRSDIPRYALAGAPVTLMI FPDYGTERTMGMHPONOCGYLGSVNAQ  
 AEKPEPNRLILHGFLEDEVHPAHTSILLSFLVRAGKPYDLIYPOERHSINVPES  
 GEHEHLHLHLOENIGSRIMALKVI"

BASE COUNT 1313 a 918 c 946 g 1358 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 5,27e-230 Length: 4535  
 Score: 2422.00 Matches: 465  
 Percent Similarity: 82.30% Conservative: 0  
 Best Local Similarity: 82.30% Mismatches: 0  
 Query Match: 97.58% Indels: 100  
 DB: 9 Gaps: 1

US-10-070-464-5 (1-465) x BCC40203 (1-4535)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 2580 ACAGGTACAGCAAAATCCCTAAAGTCACTTTAAGATGCAGAAATATGATGATGCTGAA 2639  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 2640 GGAAGATCATAGATGATCATAGATAGAGAACTAATTCACCTTTGAGATTCTATTGAA 2699  
 QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 2700 GGAGTTGAATATATGTCAGAGCTGATGAGACTCCTGAGGAAATATAGCTTGCTCATC 2759  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 2760 CTACTAGATGCTCCCTCCAGACTCGCTACAGATAGGTGATCTCACCTGAATTTATTATC 2819  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 2820 CCAATAGAGATGATGATGATGATGAAAGCAGAGACTGATGATGATGATGATGATGATG 2879  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 2880 ACGGCACATTAATCTATGAAAGAAACAACACATCTGATTAATATCATGACATCTTT 2939  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 2940 CATGTTTTTCCCAAGTCAGAGAGGAAATGAGTTATTTTTCCTGGAATGCAAA 2999  
 QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 DB 3000 ACAGGTTCCCTCATTTATACAAATTTATCATCTATTTTAAAGGAAAGCAATATTAACA 3059  
 QY 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 DB 3060 TCCAGTGTGGCTGCTGCTGCTCAAGTATTCATTCCTCATCAAGAGAGAGATGACA 3119  
 QY 181 IleThrSerGlyGluTyrPheGluValIleLeuGlyValArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 3120 ATTAACCATGAGTAATGGAGAGTCTTGCCGCGCATGATCTTAATATCAAGTTGATGAA 3179  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyr 220  
 DB 3180 GTGAGAAAGCTGATATATTTGAAAGGACCAAAAGCTCCCTTTAAGACATCACTGTAC 3229  
 QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 3240 GTAGTCTAGTTACGTAATCTGAGAGGTGACAAAGCTGACTGACCGGCTACTCTACAT 3299  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 DB 3300 TCTTGCTGATCACTGACGACTGCTGCTTTTATTAAGTATAGTAACCAAGAAAT 3359  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 3360 CCACACTGTGTGCTCCCTTACAGCTATCAAGCTCCTGAAAGATGACCCCACTTGCAAAA 3419

QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 DB 3420 AAGGAATTTGGGCGCACCATTTGATTCAGACAGCTCTCTCTGACTATATCTCTCCA 3479  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 3480 GAAATTTCTCTTTTGAAGATCACTGATTAATTCATTTGATGAGATGCTCTACAAAGCT 3539  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 3540 CATATCTACAGCTCGAAGAAATATCTTACTGCTGCTTCAATATATGATGATGCTCTAG 3599  
 QY 340 ----- 340  
 DB 3600 GTGCGATTGATTAATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTAGCC 3659  
 QY 340 ----- 340  
 DB 3660 TCTCTAGTTATGTTGTTAGTATGATAGACAAAGGAGATCTGTACCGAGGCTTAAA 3719  
 QY 340 ----- 340  
 DB 3720 TTTGAAGCGCTTAAATATATAATGATCAATATGAAATGACGATCAGGTGAAAGA 3779  
 QY 340 ----- 340  
 DB 3780 CTCCAATATCTAGCTTCTCGATATGATTTGATGATGATGATGATGATGATGATGATG 3839  
 QY 340 ----- 340  
 DB 3840 TGGTCTATGAGAGATACCTCTCCCTGATGAGCATTAATGACAGGCTAGATATCTTAGG 3899  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 3900 GTTGCTATTTGCTGGGGCCCAAGTCACTGTGATCTTTCTATGATACAGATACAGGAA 3959  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 3960 CGTTATATGGGTCACTCTGACCAAGATGAACAGGGCTATTACTTAAGATCTGTGGCCATG 4019  
 QY 361 GlnAlaGluLysPheProSerGluProAspArgLeuLeuLeuHisGlyPheLeuAsp 400  
 DB 4020 CAAAGCAAAAGTTCCTCTGAAACCAATCGTTTACTGCTTACATGATGTTCTTGAT 4079  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 4080 GAGAAATGCTATTTGACATACATACATATATCTGAGTTTATGAGAGGCTGCAAG 4139  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 4140 CCATATATATTACAGATCTATCTCCAGAGAGACACAGCATTAAGATTCCTGAAATCGGCA 4199  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnLeuAsnLeuGlySerArgIleAla 460  
 DB 4200 GAAATATATGACTGCAATCTTTTGACCTTCAAGAAACCTTGATGATGATGATGATG 4259  
 QY 461 AlaLeuLysValIle 465  
 DB 4260 GCTTAATAAGTGATA 4274

Search completed: October 16, 2003, 00:04:57  
 Job time : 513.73 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:14:56 [Search time 341.661 Seconds  
(without alignments)  
3673.932 Million cell updates/sec]

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTKFMSKIMIDAE.....HLHYLGKISRIALKVI 465

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+g2n.model -DEV=xlh  
-Q=cg12\_1/USPTO.spool/US10070464/runat\_15102003\_113552\_24799/app\_query.faeta\_1.2652  
-DB=N-Geneseq\_19jun03 -OFMT=faeta -SUFFIX=ring -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomum2 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptlo -NOBM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10070464 @CCH 1.1 0 @runat\_15102003\_113552\_24799 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N-Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and lg derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2482	100.0	1669	22 AAC85696	Nucleotide sequenc
2	2482	100.0	2668	24 ABN59775	Novel human coding
3	2482	100.0	2842	24 ABN59774	Novel human coding
4	2446.5	98.6	4523	24 ABK83325	CDNA encoding huma
5	2441.5	98.4	2161	22 AAH15009	Human cDNA sequenc
6	2422	97.6	2643	24 AAH99935	Coding sequence of
7	2422	97.6	2649	25 ABX12255	CDNA encoding huma
8	2422	97.6	2671	24 ABK83322	CDNA encoding huma
9	2422	97.6	3106	24 ABK12892	Human protease PR
10	2422	97.6	3120	22 AAC85694	Nucleotide sequenc
11	2422	97.6	3120	24 AAD89956	Human dipeptidyl p
12	2422	97.6	3143	24 AAH99934	CDNA encoding 2195
13	2418.5	97.4	4676	24 ABK83331	CDNA encoding huma
14	2402	96.8	4829	24 ABK83327	CDNA encoding huma
15	2244.5	90.4	2510	24 AAD23843	Human protease PR
16	2219	89.4	1821	24 ABV76411	Dipeptidyl peptida
17	2149	86.6	4685	24 ABK83332	CDNA encoding huma
18	1494.5	60.2	2495	24 AAD89957	Human dipeptidyl p
19	1494.5	60.2	2617	24 ABK83323	CDNA encoding huma
20	1494.5	60.2	3024	24 AAD89954	Human dipeptidyl p
21	1494.5	60.2	3716	24 ABQ75955	Human PMWV encodin
22	1494.5	60.2	4219	24 ABK83335	CDNA encoding huma
23	1494.5	60.2	4302	24 ABK83333	CDNA encoding huma
24	1488.5	60.0	3287	24 AAD89955	Alternative versio
25	1445	58.2	4180	24 ABK83339	CDNA encoding huma
26	1445	58.2	4263	24 ABK83338	CDNA encoding huma
27	1422.5	57.3	2952	24 ABK69090	CDNA encoding huma
28	1418.5	57.2	2751	24 AAD89113	Murine dipeptidyl
29	1418.5	57.2	3047	24 ABK69111	CDNA encoding huma
30	1386	55.8	4309	24 ABK83328	CDNA encoding huma
31	1377.5	55.1	1197	22 AAC85695	Nucleotide sequenc
32	1342.5	54.1	4076	24 ABK83337	CDNA encoding huma
33	1342.5	54.1	4159	24 ABK83336	CDNA encoding huma
34	1296	52.2	3262	22 AA157880	Human polynucleoti
35	1293	52.1	4037	24 ABK83341	CDNA encoding huma
36	1293	52.1	4120	24 ABK83340	CDNA encoding huma
37	1289	51.9	2982	22 AA159666	Human polynucleoti
38	1276	51.4	2801	22 AA157896	Human polynucleoti
39	1188.5	47.9	2027	21 AACT7137	Human ORFX ORF2692
40	1008.5	40.6	2461	21 AACT5835	Human ORFX ORF1390
41	860.5	34.7	1083	22 AAC85697	Nucleotide sequenc
42	845	34.0	3713	23 AB110425	Drosophila melanog
43	845	34.0	3783	23 AB106641	Drosophila melanog
44	784.5	31.6	1837	24 ABK69114	DNA encoding huma
45	673.5	27.1	6225	23 AB110424	Drosophila melanog

## ALIGNMENTS

RESULT 1  
AAC85696  
ID AAC85696 standard; cDNA; 1669 BP.

AC AAC85696;  
XX 29-JUN-2001 (first entry)

DE Nucleotide sequence encoding human DPP8 318Thr-658Val+759Ala-882Ile.  
XX Human dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;  
XX dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;  
KW growth hormone deficiency; glucose level; mucosal regeneration;  
KW non-insulin dependent diabetes mellitus; glucose intolerance;  
KW immunosuppression; ss.  
XX Homo sapiens.

XX Key Location/Qualifiers  
 FH 2..1399 /\*tag= a  
 FT CDS /product= "Human DPP8 318Thr-658Val+759Ala-882Ile"  
 FT  
 XX  
 PN W0200119866-A1.  
 XX  
 PD 22-MAR-2001.  
 XX  
 PF 11-SEP-2000; 2000WO-AU01085.  
 XX  
 PR 10-SEP-1999; 99AU-0002762.  
 PR 18-FEB-2000; 2000AU-0005709.  
 XX  
 PA (UNSY ) UNIV SYDNEY.  
 XX  
 PI Abbott CA, Gorell MD;  
 XX  
 DR WPI; 2001-281520/29.  
 DR P-PSDB; AAB47189.  
 XX  
 PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving  
 PT substrates, identifying inhibitors of DPP8 catalytic activity which  
 PT have therapeutic uses, and for detecting activated T cells  
 XX  
 PS Claim 21; Page 72-74; 78pp; English.  
 XX  
 CC The sequences given in AAC05695-97 encode fragments of human  
 CC dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for  
 CC H-Ala-Pro-pNa, H-Gly-Pro-pNa and H-Arg-Pro-pNa. Therefore, it is a  
 CC prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable  
 CC of hydrolysing the peptide bond C-terminal to Pro in each of these  
 CC compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for  
 CC cleaving a substrate, and for detecting an activated T cell which  
 CC involves measuring the level of DPP8 gene expression in a T cell. The  
 CC level of DPP8 expression is detected by detecting the amount of DPP8  
 CC RNA in the cell. It is also useful for identifying a molecule capable  
 CC of inhibiting the cleavage of the substrate by DPP8. Molecules  
 CC identified as inhibiting DPP8 catalytic activity may be useful for  
 CC treating diarrhoea, growth hormone deficiency, lowering glucose levels  
 CC in non-insulin dependent diabetes mellitus and other disorders  
 CC involving glucose intolerance, enhancing mucosal regeneration and  
 CC as immunosuppressants.  
 XX  
 SQ Sequence 1669 BP; 524 A; 336 C; 336 G; 473 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1.13e-254 Length: 1669  
 Score: 2482.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-070-464-5 (1-465) x AAC05696 (1-1669)  
 QY 1 ThrGlyThralaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 2 ACAGGTACAGCAATCTTAAGTCACTTTTAAGATGTCAGAAATATATGATGAGCTGAA 61  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 62 GGAAGGATCATAGATGTCATAGATTAAGAACTTAATCACTTTGAGATTCTATTGAA 121  
 QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 DB 122 GGAATTGATATATTGCGACAGCTGGAGACTCTCTGAGGAAATATGCTGATCCTC 181  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 DB 182 CTACTAGATCGCTCCAGACTCGCTACAGATAGTGTGATCTCAGCTGAATTTATTC 241

QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 242 CCAGTAGAGATGATGTTATGGAAGCAGAGACTCATTTAGTCAGTCCGATTCGTG 301  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 302 ACCGCACTAATTTATCTATGAGAAACACAGACACTCTGGATTAATATCCATGACATCTTT 361  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 DB 362 CATGTTTTCCCAAGTCAAGAGGAAATGAGTTATTTTGGCTCTGAAATGCCAA 421  
 QY 141 ThrGlyPheArgHisLeuTyrIleTyrSerIleLeuLysGluSerLysTyrLysArg 160  
 DB 422 ACAGTTTCCGTCATTTATCAAAATTAACATCTATTATTAAGGAAACCAATATTAACGA 481  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 DB 482 TCAGTGAGGCTGCTGCTCCAGATGATTTCAAGTCTCTATCAAGAGAGATAGCA 541  
 QY 181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 DB 542 ATTACCAAGTGATGATGGAAGTTCTTGCCCGCATGATCTATATCCAAAGTTGATGAA 601  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeuTyr 220  
 DB 602 GTCAAGAGCTGTGATATTTTGAAGGCACCAAGACTCCCTTTAGAGCATCACCTGAC 661  
 QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 DB 662 GTATCTAGTATGTAATTCCTGGAGAGATGACAAAGCTGACCGTGCTACTCAT 721  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsn 260  
 DB 722 TCTTGTCGATCAGTCACTGACACTGACTTCTTTTAAGTATGATGATACCAAGAT 781  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 DB 782 CCACACGTGTGTCCCTTTCACAGCTATCAAGTCTTAAGATGACCAACTGTGAAACA 841  
 QY 281 LysGluPheThrPalaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 DB 842 AAGGAATTTGGGCGACCATTTTGATTCAGCAGAGTCTCTTCTGACTATACTCTCCA 901  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 902 GAATTTTCTCTTTGAAGTACTGATGATTTATGATGATGATGATGATGATGATGATGAT 961  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 962 CATGATCTACAGCTGGAAAGAAATATCTCATGCTGTGCATATATGATGATGATGATGAT 1021  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 1022 GTTCTATTGCTGGGCGCCAGTCACTGTGATCTCTTATGATGATGATGATGATGATGAT 1081  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 1082 CGTATATGGGTACCTCTGACCAAGATGAACAGGCTATTACTTAAGATCTGTGGCATG 1141  
 QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 DB 1142 CAAGCAAAAAATTCCTCTGAAACAAATCGTTATGCTCTTACATGATGATGATGATGAT 1201  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 1202 GAGATATCCATTTTGCACATACAGATATATATGATGATTTTATAGTGGCTGGAAAG 1261  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 DB 1262 CCATATGATTTTACAGATCTATCTCCAGAGACACACAGCATTAAGATGTTCCGTAATCGGA 1321  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460

DB 1322 GAACATTATGAAGTGCATCTTTGGCACTACCTTCAAGAAAACCTTGATCAGTATTCCT 1381  
QY 461 AAlaLeuLeuValIle 465  
DB 1382 GCTCTAAAGTGATA.1396

## RESULT 2

ABNS9775

ID ABNS9775 standard; cDNA; 2668 BP.

AC ABNS9775;

DT 28-JUN-2002 (first entry)

DE Novel human coding sequence SEQ ID NO: 186.

XX Human; antianemic; vulnerrary; antiinflammatory; immunomodulator;  
XX antifertility; cerebroprotective; cytosstatic; rheumatic; gene therapy;  
XX neuroprotective; antiparkinsonian; protein therapy; EST;  
XX expressed sequence tag; gene; ss.

OS Homo sapiens.

PN WO200222660-A2.

PD 21-MAR-2002.

PF 10-SEP-2001; 2001WO-US26015.

PR 11-SEP-2000; 2000US-0659671.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX MPI; 2002-292408/33.

DR P-PSDB; ABB97362.

PT An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -

PS Claim 1; SEQ ID NO 186; 509pp; English.

XX The present invention provides the protein and coding sequences of 444  
XX novel human proteins. These were isolated from expressed sequences tags  
XX (ESTs). They can be used to stimulate cell growth, to regulate  
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
XX e.g. in burn treatment, to regulate the immune system e.g. to treat  
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat  
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat  
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions  
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
XX Parkinson's disease. The present sequence is a coding sequence of the  
XX invention.

SQ Sequence 2668 BP; 796 A; 564 C; 592 G; 716 T; 0 other;

## Alignment Scores:

Pred. No.: 2,3e-254 Length: 2668  
Score: 2482.00 Matches: 465  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-10-070-464-5 (1-465) x ABNS9775 (1-2668)

QY 1 ThrglyThrAlaAspProLysValAlaThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 1011 ACGGTACAGAAATCCCTAAAGTCACTTTAAGATGTCAGAAATATGATGATGCTGAA 1070

QY 21 GIYArgIleIleAspValIleAspLysGluLeuIleGluProPheGluIleLeuPheGlu 40  
DB 1071 GGAAGATCATGATGATTCATAGTAAGAACTAATTCACCTTTAGATTCTAATTGAA 1130  
QY 41 GIYValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
DB 1131 GGAAGTTGAATATATTCAGAGCTGGATGAGTCTCCTGAGAGAAATATGCTTGCTCAGC 1190  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1191 CTAAGTATGATGCTCCAGACCTGCTCAGATAGTGTATGATCCACCTAATTAATTATTC 1250  
QY 81 ProValGluAspAspValMetGluArgGluArgLeuIleGluSerValProAspSerVal 100  
DB 1251 CCGATGAAAGATGATGATTTGAAAGGACAGACATCATTAAGTCACTGCTGATTCGTG 1310  
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
DB 1311 ACGCCACTAATATATGATGAAACAAACAGACATCTGGATTAATATCATGATCATCTTT 1370  
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1371 CATGTTTTTCCCAAGTCACAGAGAAATTTGATTTTGGCTCTGATGATGCAAA 1430  
QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
DB 1431 ACGGTTTCGTCATTTATTAACAATTCATCTAATTTAAAGAAAGCAAAATTAACA 1490  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
DB 1491 TCCAGTGGGCGCTCGCTGCTCAAGTATTTCAAGTGTCCATCAAGAGAGATAGCA 1550  
QY 181 ILeThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1551 ATTACCAAGTGGTGAATGGAGTTCTTCCGCGCATGATCTAATATCAAGATTGATGAA 1610  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLysTyr 220  
DB 1611 GTCAGAGGCTGTATATTTGAGGACCAAGAGCTCCCTTTAGAGATCACTGTAC 1670  
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1671 GTAAGTCACTTAACTTAACTCGAGAGAGTCAAGGCTGACGACCTGTGCTACAT 1730  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
DB 1731 TCTTCTCATCAGACGACACACTGTGATCTTTAATAGTAAGTATAGTAACAGAAAGAT 1790  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1791 CCACCTGTGTCTCCCTTACAACTATCAAGTCTCTGAAGAGACCCAACTTGCAAAACA 1850  
QY 281 LysGluPheThrPalaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 1851 AAGGAATTTGGGCCACATTTTGGATTCACAGGTCCTCTTCGACATATCTCTCCA 1910  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 1911 GAATTTTCTCTTTGAAAGTACTAGATTATCATGTATGAGATCCTCAACAGCT 1970  
QY 321 HisAspLeuGlnProGlyLysLeuTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
DB 1971 CATGATCTACAGCCGGAAGAAATATCTTCTGTGCTGATCATATATGATGATGCTCAG 2030  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360  
DB 2031 GTTGCTATTTGCTGGGGGCCAGTCACTGTGATCTTTATGATACGATACGAGAA 2090  
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
DB 2091 CGTTATATGGGTCAACCTGACCAAGATGAACAGGGCTATTAAGATCTGTGGCCATG 2150  
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400



Db 2151 CAAGCGAAAGAGTCCCTCTGAACCAATCGTTTACTGCTTACATGGTTCCCGAT 2210  
 Qy 401 GUAaValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2211 GAGAACTCATTTTCCACATACCAAGATATTCTAGATTTTATGAGGGCTGCAAG 2270  
 Qy 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGlnSerGly 440  
 Db 2271 CCATATGATTTACGATCTCTCTCAAGAGACACAGCATTAAGATTCCTGAATCGGA 2330  
 Qy 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 Db 2331 GAACATTATGAACTGCACTTTTGACACTTCAAGAAACCTTGATCAGATTGGCT 2390  
 Qy 461 AlaLeuLysValIle 465  
 Db 2391 GCTCTAAAGTGATA 2405  
 RESULT 3  
 ABNS9774 ID ABNS9774 standard; cDNA; 2842 BP.  
 AC ABNS9774;  
 DT 28-JUN-2002 (first entry)  
 DE Novel human coding sequence SEQ ID NO: 185.  
 KM Human; antihaemic; vulnereary; antiinflammatory; immunomodulator;  
 KM antifertility; cerebroprotective; cytoskeletal; rheumatic; gene therapy;  
 KM neuroprotective; antiparkinsonian; protein therapy; EST;  
 KM expressed sequence tag; gene; ss.  
 OS Homo sapiens.  
 OS MO200222660-A2.  
 PN 21-MAR-2002.  
 PF 10-SEP-2001; 2001WO-US26015.  
 PR 11-SEP-2000; 2000US-0659671.  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QH, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
 DR WPI: 2002-292408/33.  
 DR P-PSDB; ABB97361.  
 PT An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis -  
 PS Claim 1; SEQ ID NO 185; 509pp; English.  
 XX The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a coding sequence of the  
 CC invention.  
 SQ Sequence 2842 BP; 857 A; 592 C; 635 G; 758 T; 0 other;  
 Alignment Scores: 2.53e-254 Length: 2842  
 Pred. No.:

Score: 2482.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Gaps: 0  
 DB: 24  
 US-10-070-464-5 (1-465) x ABNS9774 (1-2842)  
 Qy 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 1185 ACAGGTACAGCAAAATCTAAAGTCATCTTTAAGATCTCGAAATATATGATTGATCGTGA 1244  
 Qy 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 1245 GGAAGGATCATGATGTCATATGATTAAGAACTCAATCACTTTGAGATTCTATTGAA 1304  
 Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrPheThrProGlnGlyLysTyrAlaTyrSerIle 60  
 Db 1305 GAGATTGAATATATATGCGACAGCTGATGACTCTGAGGAGAAATATGCTTGCTCATC 1364  
 Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1365 CTACTAGATCGCTCCAGACTCGCTACAGATAGTGTGATCTCACCTGAATTATTATTC 1424  
 Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1425 CCAGTGAAGATGATGTTATGAAAAGGCAGACATCATGAGTCAGTCCCTGATTCGTG 1484  
 Qy 101 ThrProLeuIleIleTyrGluGluThrPheAspIleThrIleAsnIleHisAspIlePhe 120  
 Db 1485 AGCCACATAATATCTATATGAAGAACACAGACATCTGATTAATATCATGACATCTTT 1544  
 Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 Db 1545 CATGTTTTCCCAAGTACGAAAGAGAAATGAGTTATTTCCTCGATGATCGAA 1604  
 Qy 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 Db 1605 ACAGGTTTCCGTCATTTATTAACAAATATACATCTATTTTAAGCAAAATATTAACGA 1664  
 Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 Db 1665 TCCAGTGTGGGCTGCTCTCCAAAGTATTCAAAGTCTCATCAAGAGAGATAGCA 1724  
 Qy 181 IleThrSerGlyGluTyrPheGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 Db 1725 ATTACAGATGGATGGAAGTCTTGGCCGGCATGATCTTAATATCAAGTGTATGAA 1784  
 Qy 201 ValArgArgLeuValIleTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220  
 Db 1785 GTCAAGAGGCTGTATATTTGAAAGCACCAAAAGCTCCCTTTAGAGCATCACCTGTAC 1844  
 Qy 221 ValIleSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1845 GTAGTATGATTACGTAAATCTCGAAGGTGACAAAGCTGACCGTGGCTATCACT 1904  
 Qy 241 SerCysValLeuSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 Db 1905 TCTTGCTGATCAGTCAGACAGTACTGATCTTTATTAAGTATATATGTAACCAAGAAAT 1964  
 Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 Db 1965 CCACACTGTGTCTCCCTTTACAAAGCTATCAACTCTCGAAGAGACCAACTTGCAAAACA 2024  
 Qy 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2025 AAGGAATTTGGGCCACATTTTGAATTCAGAGATCTCTTCCTGACATATCTCTCCA 2084  
 Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 Db 2085 GAATTTTCTCTTTGAAGATGACTAGATGATTTACATGTATGAGATGCTCAAGCCT 2144  
 Qy 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 340



```
DB 2145 CATGATTTACAGCTGGAAGAAATATCTACTGTGCTTTCATATATGGTGCTCCAG 2204
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGly 360
DB 2205 GTTGCTATTCTGGGGCCCACTGCTGAGTCTTCTATGATACAGATACACGGAA 2264
QY 361 ArgTyrMetCylHisPheAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380
DB 2265 CGTTATATGGGTACCCCTGACCAAGATGAACAGGGCTATTACTTAGATCTGTGGCCATG 2324
QY 381 GlnAlaGluPhePheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400
DB 2325 CAAGCAGAAAAGTTCCTCCCTGAAACCAATCGTTTACTGCTTCAACATGGTTCCGGAT 2384
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420
DB 2385 GAGAAATGTCATTTTGCACATACAGATATATCTAGTGTTTTAAAGAGGGCTGGAAG 2444
QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440
DB 2445 CCATATGATTTTACAGATCTATCTCCAGAGAGACACAGCATAAAGAGTTCTGAAATCGGGA 2504
QY 441 GlnHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460
DB 2505 GAAATATTATGAACTGGCATCTTTTGCACCTTCAAGAAAACCTTGGATCAGCATTTGCT 2564
QY 461 AlaLeuLysValIle 465
DB 2565 GCTCTAAAAGTGATA 2579
RESULT 4
ABK83325
ID ABK83325 standard; cDNA; 4523 BP.
XX
AC ABK83325;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPP-1 splice variant #1.
XX
DE Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
DR WPI; 2002-44178/47.
XX
DR P-PSDB; ABG61594.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
PS disclosure; Page 61-62; 113p; English.
XX
```

```
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPPs). The dipeptidyl peptidase IV-related proteins (DPP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenias), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPP proteins.
XX
SQ Sequence 4523 BP; 1384 A; 828 C; 940 G; 1371 T; 0 other;
Alignment Scores:
Pred. No.: 0 3,16e-250 Length: 4523
Score: 2446.50 Matches: 463
Percent Similarity: 99.57% Conservative: 0
Best Local Similarity: 99.57% Mismatches: 1
Query Match: 98.57% Indels: 2
DB: 24 Gaps: 1
US-10-070-464-5 (1-465) x ABK83325 (1-4523)
QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
DB 1165 ACAGATACAGCAATCTTAATGATCACTTTTAAATGTCACGAATAATATGATGATGCTGAA 1224
QY 21 GlyArgIleIleLeuPheValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40
DB 1225 GGAAGGATCATTAATGATGATGATTAAGAACTAATTCACCTTTTGAATCTTATTTGAA 1284
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60
DB 1285 GGAATGATATATTTTGCAGAGCTGGATGATGATCTCGAGGAAATATGCTTGTTCATC 1344
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
DB 1345 CTACTAGATGCTCCCGACGACTCGCTACAGATAGTGTGATTCACCTGAAATATTTATC 1404
QY 81 ProValGluAspAspValMetGluValArgGlnArgLeuIleGluSerValProAspSerVal 100
DB 1405 CCAGTGAAGATATATGTTAAGAAAGCAGACGATCAATGACAGCTGATTCGTG 1464
QY 101 ThrProLeuIleIleTyrGluGluTyrThrAspIleTyrIleAsnIleHisAspIlePhe 120
DB 1465 ACCGCACTAATTTCTATGAGAAACACAGACATCTGATTAATATCCATGACATCTTT 1524
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140
DB 1525 CATGTTTTCGCCCAAGTACGAAAGAGAAATGAGTTTATTTTGGCTCTGAAATCGAA 1584
QY 141 ThrGlyPheArgHisLeuTyrIleTyrSerIleLeuLysGluSerLysTyrLysArg 160
DB 1585 ACAGGTTTCCGTCAATTTATCAAAATTTACATCTTATTAAGAAAGCAAAATTAACGA 1644
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180
DB 1645 TCCAGTGGTGGGCTGCTGCTCCAGAGTATTAAGTGCTCTATCAAAAGAGAGATAGCA 1704
QY 181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200
DB 1705 ATTAACAGAGGTGAATGGGAAGTCTTGGCCGCAAGATCTAATATCCAAAGTTGATGAA 1764
QY 201 ValArgArgLeuValIleTyrPheGluGlyThrLysAspSerProLeuGluHisIleLysTyr 220
DB 1765 GTCAAGAGCGGTGATATTTTGAAGGACCAAGACCTCCCTTTAGGATCATCCTGTAC 1824
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240
```

Db 1825 GTAGTCAGTACGTAATTCCTGGAGAGGTGACAAAGCTGACTGACCGTGCTACTACAT 1884  
 Qy 241 SerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGlnLysAsn 260  
 Db 1885 TCTTGCTGCATCAGTCACACACCTGTCATCTTTTAACTAAGTAACTAAGTAACTAAGTAA 1944  
 Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 Db 1945 CCACAGCTGTGTCTCCCTTACACAGCTATCAAGTCTCTGAAGATGACCCAACTGGCAAAACA 2004  
 Qy 281 LysGluPheThrPalaThrIleLeuAspSerIleGlyPheLysProAspTyrThrProPro 300  
 Db 2005 AAGGAATTTGGCCACCATTTTGGATTCAGAGGTCTCTTCGACTTAACTCTCTCA 2064  
 Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 Db 2065 GAAATTTTCTCTTTGAAAGTAACTACTGATTTTAACTTATGATGATGATGATGATGATGAT 2124  
 Qy 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 Db 2125 CATGATCTACAGCTCGGAAAGAAATATCTACTGCTGCTTATGATGATGATGATGATGATGAT 2180  
 Qy 341 ValAlaIleAlaGlyAlaProValThrLeuThrPheTyrAspThrGlyTyrThrGlu 360  
 Db 2181 GTGCTATGCTGGGCCCCCAGTCTCTGTGATCTCTATGATGATGATGATGATGATGATGATGAT 2240  
 Qy 361 ArgTyrMetGlyHisProAspGlnAsnGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2241 CGTTATATGGCTCAACCTGACACAGATGACAGGCTTATGATGATGATGATGATGATGATGATGAT 2300  
 Qy 381 GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAsp 400  
 Db 2301 CAACAGAGAAAGTTCCTCTGAAACCAATGGTTTACTGCTTACATGATGATGATGATGATGATGAT 2360  
 Qy 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2361 GAGAAATGCTCTTTTGCACATACCGATATATTAAGTATTTTAAAGAGGCTGGAAAG 2420  
 Qy 421 ProTyrAspLeuGlnIleTyrProGlnLysArgHisSerIleArgValProGluSerGly 440  
 Db 2421 CCATATGATTTTACAGATCTATCTCTCAGAGAGACACATATAGAGTTCCTGAATCCGGA 2480  
 Qy 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnLysLeuLeuGlySerArgIleAla 460  
 Db 2481 GAAATATTAAGACGTCATCTTTTGCATCACTTCAAGAAACCTTGATGATGATGATGATGATGAT 2540  
 Qy 461 AlaLeuLysValIle 465  
 Db 2541 GCTCTAAAGTGA 2555  
 RESULT 5  
 AAH15009  
 ID AAH15009 standard; cDNA; 2161 BP.  
 XX  
 AC AAH15009;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12963.  
 XX  
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 12963; 2537bp + CD ROM; English.  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to a  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and  
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95993 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 3,54e-250 Length: 2161  
 Score: 2441.50 Matches: 464  
 Percent Similarity: 90.27% Conservative: 0  
 Best Local Similarity: 90.27% Mismatches: 1  
 Query Match: 98.37% Indels: 49  
 DB: 22 Gaps: 1  
 US-10-070-464-5 (1-465) X AAH15009 (1-2161)  
 Qy 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 370 ACAGGTACAGCAAAATCCTAAAGTCACTTTAAAGATGTCAGAAATTAAGATGATGATGATGAA 429  
 Qy 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 430 GGAAGCATCATAGTGTCTATGATTAAGAACTAATTTCACTTTTGAGATTCTATTGAA 489  
 Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 Db 490 GGAATTAATTAATATGTCAGAGCTGATGATGATCTCTGAGGAAATATGCTGTGTCATC 549  
 Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 550 CTACTTAATGCTCTCCAGACTCGCTTACAGATGTTGATCTCACTGAAATTAATTAATC 609  
 Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 610 CCACTTAAGATGATGTTTAAAGGCAAGACTCATTAAGTACAGCTGATTTCTGTG 669  
 Qy 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120

```

Db      670  ACCGCACTAATATCTATGAAAGAAACAAGACATCGATTAATATCCATGACATCTTT 729
Qy      121  HieValPheProGlnSerHisgluGluGluileGluPheIlePheAlaSerGluCysLys 140
Db      730  CATGTTTTTCCCAAGTACAGAAAGAGAAATTGAGTTATTTTGGCTCTGGAATGCAAA 789
Qy      141  ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160
Db      790  ACGAGTTTCCGCTCTTTATACAAATTAATCATCTATTATTAAGAAAGCAATATAAGCA 849
Qy      161  SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluIleAla 180
Db      850  TCCAGTGAGTGGGCTGCTGCTCCAAAGATTTCAAGTCTCTATCAAAAGAGAAATGCA 909
Qy      181  IleThrSerGlyLysIleTyrGluValLeuGlyArgHisgluSerAsnIleGluValAspGlu 200
Db      910  ATTACCGAGTGTGATGAGGAAGTCTTGGCCGAGATGATCTAATATCCAAAGTTGAGAA 969
Qy      201  ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220
Db      970  GTCCAGAAAGCTGTATATTTTGAAGGCAACCAAGACTCCCTTTAGAGCATCACTGTAC 1029
Qy      221  ValValSerTyrValAsnProGlyGlyValThrArgLeuThrAspArgGlyTyrSerHis 240
Db      1030  GTAGTCAGTACGTAATCTCGAGAGGTGCAAGGCTGACCTGACCGTAGCTACTCAT 1089
Qy      241  SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260
Db      1090  TCTTGCTGCATCGATCGACACTGTGACTCTTTATAAGTAAGTAACTAATCAAGAAAGAT 1149
Qy      261  ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280
Db      1150  CCACACGTGCTGCTCCCTTTCAAGCTATCAAGTCTCTAAGATGACCCAACTGCAAAACA 1209
Qy      281  LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300
Db      1210  AAGGAATTTTGGGCCCACTATTTGGATTTCAGAGGTCTCTCTGACTACTACTCTCCCA 1269
Qy      301  GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320
Db      1270  GAAATTTTCTCTTTGAAATACACTGAGATTATGATGAGAGCTCTCAAGACCT 1329
Qy      321  HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340
Db      1330  CAGGATCTACAGCTCGAAAGAAATATCTACTGTGCTGTTCATATATGTGTCTCTCAG 1389
Qy      340  ----- 340
Db      1390  GTGCAGTTGGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCCCTAGCC 1449
Qy      340  ----- 340
Db      1450  TCTCTAGGTTATGTGTTGATGATAGACACAGGGGATCTGTGACCGAGGGCTTAA 1509
Qy      341  -----ValAlaIleAlaGlyAlaProValThrLeuTyr 351
Db      1510  TTTGAAGGGCCCTTAATATTAATATGATGCTGATTCCTGGGCCCAAGTCACTCTGTG 1569
Qy      352  IlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGluGln 371
Db      1570  ATCTTCTATGATACAGGATACAGGAACGTTATATGGGTCAACCTGACCGAATGAACAG 1629
Qy      372  GlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPheProSerGluProAsnArg 391
Db      1630  GGGTATTACTTAAAGATCTGTGGCAATGCAAGCAAGAAAGTTCCCTCGAACAACAAATCGT 1689
Qy      392  LeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeu 411
Db      1690  TTAAGTCTCTTACATGAGTTTCCCGATGAGAAATGTCATTTTGCAATACACATATATTA 1749
Qy      412  LeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTyrProGlnGluArg 431

```

```

Db      1750  CTGAGTTTTTTATGAGAGGCTGGAAAGCATATGATTTACATCTATCTCAGAGAGA 1809
Qy      432  HisSerIleArgValPProGluSerGlyGluHisTyrGluLeuHisLeuHisTyrLeu 451
Db      1810  CACAGCATTAAGAGTTCCTGATCGGAGAAATTATTAATGATCTTTTGACACTACTT 1869
Qy      452  GlnGluAsnLeuGlySerArgIleAlaIleLeuValIle 465
Db      1870  CAGGAAACCTTGATCAGTATTTCTGCTCTTAAGAGATTA 1911

RESULT 6
AAH99935
ID  AAH99935 standard; cDNA; 2643 BP.
XX
AC  AAH99935;
XX
DT  12-APR-2002 (first entry)
XX
DE  Coding sequence of 21953 human prolly1 oligopeptidase.
XX
KM  21953 prolly1 oligopeptidase; antibody; proline; endopeptidase;
KM  cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KM  neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KM  antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KM  diabetes mellitus; arthritis; multiple sclerosis; asthma;
KM  Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
OS  Homo sapiens.
XX
PN  WO200179473-A2.
XX
PD  25-OCT-2001.
XX
PF  11-APR-2001; 2001WO-US40483.
XX
PR  18-APR-2000; 2000US-197508P.
XX
PA  (MILL-) MILLENNIUM PHARM INC.
XX
PT  Meyers RA, Williamson M;
XX
DR  WPI; 2002-034353/04.
XX
PT  P-PSDB; AAG78415.
XX
PT  New polypeptides 21953, member of human prolly1 oligopeptidase family,
PT  useful as diagnostic targets and therapeutic agents for controlling
PT  cancer, lymphoma and leukemia -
XX
PS  Claim 2; Page; 121pp; English.
XX
CC  This invention relates to an isolated 21953 human prolly1
CC  oligopeptidase, which is cytosolic, antidiabetic, antiarthritic,
CC  neuroprotective, antithyroid, dermatological, antiproliferative,
CC  antiasthmatic, ophthalmological, antiinflammatory, neotropic,
CC  antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC  antineural, cardiant, antihypertensive, anorectic and
CC  metabolic in its action. Uses include gene therapy, expression or
CC  activity of 21953 protein modulator, it is useful for identifying a
CC  compound which binds to it and can be used in preventing, treating
CC  or detecting a cellular proliferative or differentiative disorder.
CC  The 21953 molecules can act as novel diagnostic targets and therapeutic
CC  agents for controlling disorders associated with the aberrant activity
CC  or degradation of peptide hormones e.g., disorders associated with cell
CC  differentiation and proliferation such as cancer, immune function,
CC  reproductive, neurological and cardiovascular function. The 21953
CC  molecules are thus useful for treating and preventing cellular
CC  proliferative and differentiative disorders, haematopoietic neoplastic
CC  disorders, immune disorders such as autoimmune diseases, diabetes
CC  mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC  neuronal disorders, demyelinating diseases, vascular diseases and
CC  metabolism or pain disorders. This sequence represents the cDNA
CC  encoding sequence of 21953 human prolly1 oligopeptidase. This
CC  sequence represents the coding sequence of 21953, being the sequence

```

CC in between the start and the stop codon of the sequence represented in  
 CC AAH99934.

SQ Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;

# Alignment Scores:

Pred. No.:	5.8e-248	Length:	2643
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	82.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	24	Gaps:	1

US-10-070-464-5 (1-465) x AAH99935 (1-2643)

QY	1	ThG1yThR1aAenProlYsValThrPhelysMetSerGluileMetileAspAlaGlu	20
DB	949	ACAGGTACAGAAATCTTAAGTCACTTTTAAGTGCAGAAATTAAGTGAAGTGCAGAA	1008
QY	21	GIYAgt1le1leAspVal1leAspLyGluLeuileGlnProPheGlu1leLeuPheGlu	40
DB	1009	GGAAGATCATAGATGATCATAGATAGAACTAATCACTTTTGAGATTCTATTGAA	1068
QY	41	GIYVal1GluYr1le1leAspAlaG1YTrPheProGluGluYsYr1le1leAspSer1le	60
DB	1069	GGAAGTAAATATATATGCGAGAGCTGAGTGAATCTCTGAGGAAATATGCTGATCATC	1128
QY	61	LeuLeuAspArgSerGlnThrArgLeuGln1leValleuileSerProGluLeuPhe1le	80
DB	1129	CTACTAGATGCTGCCAGACCTGCTGAGTATGATGATCTCACTGAAATTTATTATC	1188
QY	81	ProVal1GluAspAspVal1MetGluArgGlnArgLeuileGluSerVal1ProAspSerVal1	100
DB	1189	CCAGTAAAGATATGATGATATGAGAAAGCAGAGACTCATGATGATGCTGATTCGTG	1248
QY	101	ThrProleuile1leYrGluGluThrThrasp1le1leAsp1le1leAsp1le1lePhe	120
DB	1249	ACCGCAATATATATATGAGAAAGCAACAGATCTGAGTAAATATCCATGACATCTTT	1308
QY	121	HisValPheProGlnSerHisGluGluGlu1leGluPhe1lePhe1leAspSerGluCysLeys	140
DB	1309	CAGTATTTTCCCAAGTCAAGAGAGAAATGATTTATTTTGGCTCTGAAATGCAAA	1368
QY	141	ThG1yPheArgHisLeuYr1leYs1leThSer1leLeuYsGluSerLyYr1leYsArg	160
DB	1369	ACAGGTTTCCGCTATTTATCAAAATTAATCTATTATTAAGGAAACAAATATTAACGA	1428
QY	161	SetSerG1yG1leuProAlaProSerAspPheLysCysProileYsGluGlu1leAla	180
DB	1429	TCCAGTGGTGGCTGCTGCTCCAGATTTTCAAGTGTCTCTATCAAGAGAGATGCA	1488
QY	181	1leThSerG1yGluTrpGluVal1leuG1yArgHisG1ySerAsn1leGlnVal1leAspGlu	200
DB	1489	ATTACAGTGGTGAATGGAGAGTTCTTGCCGGCATGATCTAATATCCAAAGTTGATGAA	1548
QY	201	Val1ArgArgLeuVal1YrPheGluGluYrThrLeAspSerProleuGluHisLeuYr	220
DB	1549	GTCAGAGGCTGTATATTTTGAAGGCAACAAAGATCCCTTTAGACATCACTGTAC	1608
QY	221	Val1ValSerYrVal1AenProG1yGluVal1ThrArgLeuThraspArgG1yYrSerHis	240
DB	1609	GTAAGTCAATATCTTAATCTGAGAGGTACAAAGCTGACCTGACCTGATCTCAT	1668
QY	241	SetCysCys1leSerGlnHisCysAspPhePhe1leSerLyYrSerAsnGlnLyAsn	260
DB	1669	TCTTGCTGATCATGACAGCTGATGATCTTTATATAGTAAATATGATTAACAGAGAAAT	1728
QY	261	ProHisCysVal1SerLeuYrYsLeuSerSerProGluAspAspProThrCysLeysThr	280
DB	1729	CCACACGTGTGTCTTTCAAGCTATCAAGTCTGAAATATACCCAACTTCCAAACAA	1788
QY	281	LyGluPheTrpAlaThr1leLeuAspSerAlaG1yProleuProAspYrThrProPro	300

DB	1789	AAGAAATTTGGGCGACCAATTTGGATTCAGCAGGTCCTTCTCTGATATCTCTCCA	1848
QY	301	Glu1lePheSerPheGluSerThrThG1yPheThrLeuYrG1yMetLeuYrLeysPro	320
DB	1849	GAATTTTCTCTTTTGAATATCTACTGATTTTATCATTTGATAGGAGCTCTACAACTT	1908
QY	321	HisAspLeuGlnProG1yLyYsYrProThrVal1leuPhe1leYrG1yG1yProGln	340
DB	1909	CATGATCTACAGCTGGAAGAAATATCTACTGTGTCTGTTCAATATATGTTGCTCTCAG	1968
QY	340	-----	340
DB	1969	GTCAGATTGGTGAATATATCGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTTAGCC	2028
QY	340	-----	340
DB	2029	TCTCTAGATTATGTTGTTAGTATAGACAAAGGGGATCTGTACACGAGGCTTAA	2088
QY	340	-----	340
DB	2089	TTTGAAGCGCTTAAATATATAATGGTCAATAGAAATGACGATCAGTGAAGGA	2148
QY	340	-----	340
DB	2149	CTCCAAATATCTAGCTTCTGATATGATTTTCATTGACTAGATCTGTGGGATCCACGC	2208
QY	340	-----	340
DB	2209	TGATCCATAGAGGATACCTCTCCATGATGATTAAGCAGAGTCAATATCTTCAGG	2268
QY	341	Val1Ala1le1leAlaPheProVal1ThrLeuTrp1lePheYrAspThrG1yYrThGlu	360
DB	2269	GTTGCTATTGCTGGGCGCCAGTCACTGTGTGATCTTATGATATCAGATATCACGGA	2328
QY	361	ArgTrpMetG1yHisProAspGlnAsnGluGlnG1yYrTrpYrLeuG1ySerVal1Amet	380
DB	2329	CGTTATATGGGTACCTGACCAAGATGAACAGAGGCTATTACTTACGATCTGTGGCCATG	2388
QY	381	GlnAlaG1yLysPhePheProSerGluProAsnArgLeuLeuLeuHisG1yPheLeuAsp	400
DB	2389	CAGCAGAAAGTTCCTCCCTGTAACCAATCGTTTACTGCTCTTACATGGTTTCGTGAT	2448
QY	401	GluAsnVal1lePheAlaHisThrSer1leLeuLeuSerPheLeuVal1ArgAlaG1yLys	420
DB	2449	GAGAAATGCTCATTTTGCATACATACAGTATATTTACTAGTTTATGAGAGGCTGGAAG	2508
QY	421	ProTrpAspLeuGln1leYrProGlnGluArgHisSer1leArgVal1ProGluSerG1y	440
DB	2509	CCATATGATTTTACAGATCTATCTCTCAGAGAGACAGCATTAAGATCTCTGAATGGGA	2568
QY	441	GluHisYrG1yLysLeuHisLeuLeuHisYrLeuGlnGluAsnLeuG1ySerArg1leAla	460
DB	2569	GAACATTAAGAACTGATCTTTTGGACATACCTTCAAGAAACCTTGATCAGCTATTCCT	2628
QY	461	AlaLeuLysVal1le 465	
DB	2629	GCTCTAAAGTGATTA 2643	

RESULT 7  
 ABX12255  
 ID ABX12255 standard; cDNA; 2649 BP.  
 XX ABX12255;  
 AC  
 XX  
 DT 19-MAY-2003 (first entry)  
 XX  
 DE cDNA encoding human serine protease H1PHM46.  
 XX Human; ss; gene; H1PHM46; serine protease; gene therapy; osteoarthritis;  
 KW serine protease activity modulation; dipeptidyl peptidase activity;  
 KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;  
 KW Alzheimer's disease; paraneuronal palsy; Huntington's disease;  
 KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;

KM irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;  
 KM haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;  
 KM colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;  
 KM multiple sclerosis; chromosome 15q21-q22.

OS Homo sapiens.

Key Location/Qualifiers  
 FT CDS 1..2649  
 FT /tag= a  
 FT /product= "HIPHUM46"  
 FT /note= "Serine protease"

GB2374869-A.

30-OCT-2002.

22-JAN-2002; 2002GB-0001404.

23-JAN-2001; 2001GB-0001760.

(GLAXO ) GLAXO GROUP LTD.

Edbrooke MR, Lewis AP;

WPI: 2003-150703/15.

P-PSDB; ABU07720.

PT Identifying modulators of serine protease activity useful for treating  
 PT musculoskeletal diseases, by contacting cell expressing a novel serine  
 PT protease polypeptide with a compound and monitoring serine protease  
 PT activity -  
 XX  
 XX  
 PS Claim 12; Page 22-26; 38pp; English.

CC The invention relates to a method of identifying a substance that  
 CC modulates serine protease activity, comprising contacting a cell such as  
 CC a neuronal cell, lung cell, intestinal cell or a cell infected with a  
 CC virus, expressing a serine protease polypeptide (HIPHUM 46), or its  
 CC variant having dipeptidyl peptidase activity, or a serine protease  
 CC isolated from the cell with a test substance and monitoring for serine  
 CC protease activity. The method is useful for identifying a substance that  
 CC modulates serine protease activity. A modulator of the serine protease is  
 CC useful in the manufacture of a medicament for treatment or prophylaxis of  
 CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus  
 CC infection, Alzheimer's disease, paraneuronal palsy, myotonic  
 CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.  
 CC Additional disease that may be treated using modulators of the serine  
 CC protease include malabsorption syndromes, irritable bowel syndrome, lung  
 CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,  
 CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,  
 CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple  
 CC sclerosis. The present sequence represents cDNA encoding the human serine  
 CC protease HIPHUM46 which is located on chromosome 15q21-q22.

XX  
 SQ Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;

Alignment Scores:

Pred. No.:	5.82e-248	Length:	2649
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	82.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	25	Gaps:	1

US-10-070-464-5 (1-465) x ABX12255 (1-2649)

QY 1 ThG1YThrAlaAsnProLySValThrPheLyMetSerGluIleMetIleAspAlaGlu 20  
 Db 952 ACAGGTACAGCAATCCCTTAAGTCACTTTAAGATGTCAGAAATPAAGATTGAGCTGAA 1011  
 QY 21 G1YArgIleIleAspValIleAspLyGluLeuIleGlnProPheGluIleLeuPheGlu 40

Db 1012 GGAAGATCATAGATGTCATAGATPAAGCACTAAATTCACCTTTGAGATTCATTTGAA 1071  
 QY 41 G1YValGluYrIleAlaArgAlaG1YrThrProGluGluYLySValATrSerIle 60  
 Db 1072 GAGGTGAATATATTGCGAGAGCTGAGTGAAGCTCTCGAGGAAATATGCTTGCTTCATC 1131  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1132 CTACTAGATGCTCCAGACACTCGCTGAGATAGTGTGATCTACCTGAAATTAATTATTC 1191  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1192 CCAGTAGAAGATGATGTTATGAGAAAGCAGAGACTCATGTGATGCTGCTGATTCGTG 1251  
 QY 101 ThrProLeuIleIleYrGluGluThrThrAspIleThrPheIleAspIlePhe 120  
 Db 1252 AGCCACATTAATTAATCTAAGAAACACACACACTCGAATAATATCATAGACATCTTT 1311  
 QY 121 HisValPheProGlnSerHisGluGluIleGluPheIlePheAlaSerGluCysArg 140  
 Db 1312 CATGTTTTCCCAAGTCAAGCAAGAAATGAGTTATTTTGCCTCGAATGCAAA 1371  
 QY 141 ThrGlyPheArgHisLeuYrIleYrIleThrSerIleLeuYrGluSerLySValArg 160  
 Db 1372 ACAGTTTCCGTCATTAATCAAAATTAATCATCTATTAAAGCAAGCAAAATATTAACA 1431  
 QY 161 SerSerGlyGluLeuProAlaProSerAspPheCysProIleYrGluGluIleAla 180  
 Db 1432 TCCAGTGTGGGCTGCTGCTCCAGTGAATTCAGTGTCTTATCAAGAGAGAGATACCA 1491  
 QY 181 IleThrSerGlyGluTrpGluValLeuGluYrArgHisGluSerAsnIleGluValAspGlu 200  
 Db 1492 ATTACCAAGTGTGAATGGAAAGTTCTGGCCGCAATGATTAATATCAAGTTGATGA 1551  
 QY 201 ValArgArgLeuValYrPheGluGlyThrLyAspSerProLeuGluHisLeuYr 220  
 Db 1552 GTCAAGAGGCTGGATATTTTGAAGGACCAAAAGCTCCCTTAAGACATCACCTGAC 1611  
 QY 221 ValValSerYrValAsnProGluGluValThrArgLeuThrAspArgGlyYrSerHis 240  
 Db 1612 GTAGTCAAGTGAATTAATCTCGAGAGGTGCAAGGCTGACTGACCGTGTACTACAT 1671  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLySValSerAsnGluAsn 260  
 Db 1672 TCTTGTCGATCATGACACTGACTGACTCTTTAATAGTATAGTATGACCAAGCAAT 1731  
 QY 261 ProHisCysValSerLeuYrLyLeuSerSerProGluAspAspProThrCysLySThr 280  
 Db 1732 CCACAGTGTGTCTCTTACAACTATCAAGTCTGAAAGATGACCAACTTGCMAACA 1791  
 QY 281 LyGluThrPheThrPheAlaThrIleLeuAspSerAlaGlyProLeuProAspYrThrProPro 300  
 Db 1792 AAGGAATTTTGGGACACCATTTTGGATTCAGCAGAGTCTCTTCTGACTAATACCTTCCA 1851  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuYrGluMetLeuYrLyAspPro 320  
 Db 1852 GAAATTTTCTTTTGAAGTACTGACTGATTTTCAATTTATGGAATGCTCTACAGCCT 1911  
 QY 321 HisAspLeuGlnProGluYrLySValYrProThrValLeuPheIleYrGlyGluProGln 340  
 Db 1912 CATGATCTACAGCTGAGAAAGAAATATCTCACTGTGCTGTTCAATATATGGGTCTCAG 1971  
 QY 340 ----- 340  
 Db 1972 GTGCAATTGTTGAATATCGGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTAGCC 2031  
 QY 340 ----- 340  
 Db 2032 TCTCTAGATTATGTTGTTAGTATAGACAAAGGGATCTGTCAACGAGGCTTAA 2091  
 QY 340 ----- 340  
 Db 2092 TTGMAAGCGCCTTAAATATATAATGGGTCAATATGAAATGACATCAGTGAAGCA 2151  
 0

```
QY 340 ----- 340
Db 2152 CTCCAATATCTAGCTTTCGATATGATTTCACTAGATCGTGGGCATCCAGCGC 2211
QY 340 ----- 340
Db 2212 TGGTCTTATGGAGGATACCTCTCCCTGATGGCATTAAATGACAGGTCAGATATCTTCAGG 2271
QY 341 ValAlaIleAlaGlyAlaProValThirLeuTrpIlePheTyraSPThrGlyTyThrGlu 360
Db 2272 GTTGCTATTTGCTGGGGCCCAAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAA 2331
QY 361 ArGlyTrMetGlyHisProAspGlnAsnGlnGlnGlyTyTyLeuGlySerValAlaMet 380
Db 2332 CGTATATATGGGTACCCCTGACCCAGATATACAGGGCTTACTTACGATCTGTGGCCCATG 2391
QY 381 GlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400
Db 2392 CAAAGCAAAAAGTTCCCTCTGMAACCAATCGTTTACTGCTCTTACATGATGTTTCTGGAT 2451
QY 401 GlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420
Db 2452 GAGATATGTCATTTTGCATATACCAAGTATATTAAGTATTTTAAAGTGGGCTGGAAAAG 2511
QY 421 ProTyrAspLeuGlnIleTyProGlnGlnLysSerIleArgValProGlnSerGly 440
Db 2512 CCATATGATTTTACAGATCTATCTCAGAGAGACACAGCATTAAGATCTCGAAATCTGGA 2571
QY 441 GlnHisTyArgIleuHisLeuLeuHisTyLeuGlnGlnAsnLeuGlySerArgIleAla 460
Db 2572 GAACATTTATGAACTGCATCTTTTGCATCACTTCAAGAAAACCTTGGATCAGATTTGCT 2631
QY 461 AlaLeuLysValIle 465
Db 2632 GCTCTAAAGTGATA 2646
RESULT 8
ABK83322
ID ABK83322 standard; cDNA; 2671 BP.
XX
AC ABK83322;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPPIV related serine protease DPPP-1.
XX
KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPPP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dykinesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
WP1: 2002-444178/47.
XX
P-PSDB: ABG61591.
XX
New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
```

the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain

Claim 1; Page 53-54; 113pp; English.

The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPPIV)-related proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, psychotic and neurological disorders (e.g. anxiety, dementia, or schizophrenia), and dyskinesias. These may also be used in discovering therapeutic agents for the treatment of reproductive, inflammatory and metabolic disorders. ABK83322-ABK83343 encode human DPPP proteins.

Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;

#### Alignment Scores:

Pred. No.:	5,896-248	Length:	2671
Score:	2422.00	Matches:	465
Percent Similarity:	82.30%	Conservative:	0
Best Local Similarity:	82.30%	Mismatches:	0
Query Match:	97.58%	Indels:	100
DB:	24	Gaps:	1

US-10-070-464-5 (1-465) x ABK83322 (1-2671)

```
QY 1 ThrGlyThraIaenProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
Db 959 ACAGGTACAGCAAAATCTTAAGTACCTTTTAAGTATGACAGAAATATGATGATGCGAA 1018
QY 21 GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGluIleLeuPheGlu 40
Db 1019 GGAAGGATCATAGATGTCATATGATTAAGAACTAATTCACCTTTTGAGATCTTATTCGAA 1078
QY 41 GlyValGlnTyrrIleAlaArgAlaGlyTyrrThrProGlnGlyLysTyrrAlaTrpSerIle 60
Db 1079 GGAATTAATTAATTTGCAAGCTGATGAGCTCTCAGAGAAATATGCTGTGTCATTC 1138
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGlnLeuPheIle 80
Db 1139 CTACTAGATCGCTCCAGAGCTCGCTGACATAGTGTGATCTCAGCTGAATTTATTC 1198
QY 81 ProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerValProAspSerVal 100
Db 1199 CCACTAGAAATGATGTTATGAGAAAGCAAGACTCATTTGATGACAGCGCTGATTCGTG 1258
QY 101 ThrProLeuIleIleTyrrGlnGluTyrrThrAspIleTyrrIleAsnIleHisAspIlePhe 120
Db 1259 ACGCCACTAATTAATCTATGAGAAACAACAGACTCTGATTAATATTCATGACATCTTT 1318
QY 121 HisValPheProGlnSerHisGlnGlnGlnIleGlnPheIlePheAlaSerGluCysLys 140
Db 1319 CATGTTTTTCCCAAAAGTCAAGAAAGCAAAATGATTTATTTTGGCTCTGAAATGCAAA 1378
QY 141 ThrGlyPheArgHisLeuTyrrLysIleThrSerIleLeuLysGlnSerLysTyrrLysArg 160
Db 1379 ACAGGTTTCGTCATTTATACAAATAATTAATCTATTTTAAAGAAACCAATAATAAACGA 1438
QY 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGlnGluIleAla 180
Db 1439 TCCAGTGTGGGCTGCTCTCCAGATGATTTTCAAGTCTCTATCAAGAGAGATGACA 1498
QY 181 IleThrSerGlyGlnTyrrGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200
Db 1499 ATTACCAAGTGTGATGGAAGTTCTTGCGCGCATGATCTAATATCAAGATGAGCAA 1558
```



```

QY 201 ValArgHrgLeuValIYrPheGluGlyThrIysAspSerProLeuGluHISLeuTYr 220
DB 1559 GTGGAAGGCTGGATATTTTGAAGGACCAAGACTCCCTTTAGAGCATCACTGTAC 1618
QY 221 ValValSerTYrValAlaAnProGlyGluValThrArgLeuThrAspArgGlyTYrSerHis 240
DB 1619 GTAGTCAGTACGTAATCCCTGGAGAGGTGACAAAGGCTGATGACCTGGCTACTGCAT 1678
QY 241 SerCysAlIeSerGlnHisCysAspPhePheIleSerIYrSerAsnGlnLYAsn 260
DB 1679 TCTTGCTGCATCACTGACACTGTGACTCTTTTAAAGTAAGTATAGTAAACGAAAGAT 1738
QY 261 ProHisCysValSerLeuTYrLYsLeuSerSerProGluAspAspProThrCysLYsThr 280
DB 1739 CCACACTGTGTGCTCTTTTCAAGCTATCAAGTCTTGAAGATGACCACTGGCAAAACA 1798
QY 281 LYsGluHethrPheIleThrIleLeuAspSerAlaGlyProLeuProAspTYrThrProPro 300
DB 1799 AAGGAATTTGGGCGACCATTTGGATTGGATTCAGACGCTCTTCTGACTATACTCTCCA 1858
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTYrGlyMetLeuTYrLYsPro 320
DB 1859 GAAATTTTCTCTTTGAAAGTACTACTGATTTTCATTTGATGGATGCTCTACAAAGCT 1918
QY 321 HisAspLeuGlnProGlyLYsLYsTYrProThrValLeuPheIleTYrGlyGlyProGln 340
DB 1919 CATGATCTACAGCTGGAAAGAAATATCTGCTGCTCATATATGGTGGTCTCTCAG 1978
QY 340 ----- 340
DB 1979 GTGCACTGGTGTGAATATGATTTTAAAGAGTCAAGTATTTCCGCTTAAATACCTTACC 2038
QY 340 ----- 340
DB 2039 TCTCTAGATTATGTTGTTAGTATGATGACAAAGGGGATCTCTGACCAAGGGCTTAAA 2098
QY 340 ----- 340
DB 2099 TTTGAAGCGCCTTTAAATATAAATGGTCAATAGAAATGACATCAGTGGAAAGA 2158
QY 340 ----- 340
DB 2159 CTCCAATATCTAGCTTCGATATGATTTTCACTTGAATGCTGTGGCATCCACGCG 2218
QY 340 ----- 340
DB 2219 TGGTCTTATGAGGATACCTCTCCCTGATGAGCATTAATGACAGAGTCAGATATCTTACG 2278
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTYrAspThrGlyTYrThGlu 360
DB 2279 GTTGCTATGCTGGGGCCCGAGTCATCTGTGATCTTCTATGATACAGGATACAGGAA 2338
QY 361 ArgTYrMetGlyHisProAspGlnAsnGluGlnGlyTYrTYrLeuGlySerValAlaMet 380
DB 2339 CGTTATATGGGTACCTCTGACCAAGATGAACAGGGCTATTAAGATCTGTGGCCATG 2398
QY 381 GlnAlaGlyLeuPheProSerGluProAsnArgLeuLeuLeuHISGlyPheLeuAsp 400
DB 2399 CAAGCAGAAAAGTCCCTCTGAACCAATCGTTTACTGCTTACATGTTTCTTGAT 2458
QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLYs 420
DB 2459 GAGAAATGTCATTTTGACATACACAGATATTAAGTATTTTAAAGAGGCTGGAAAG 2518
QY 421 ProTYrAspLeuGlnIleTYrProGlnGluArgHisSerIleArgValProGluSerGly 440
DB 2519 CCATATGATTTTAAAGATCTATCTCCAGAGAAAGACACGCAATTAAGAGTCTTGATGGGA 2578
QY 441 GluHISTYrGlyLeuLeuHisLeuLeuHISTYrLeuGlnGluAsnLeuGlySerArgIleAla 460
DB 2579 GAACATTTATGAAGTCTTTTGGCACTTACCTTCAAGAAAACCTTGATCAGCTATTTGCT 2638
QY 461 AlaLeuLYsValIle 465

```

```

DB 2639 GCTTAAAGTGATA 2653
RESULT 9
ABK12892
ID ABK12892 standard; cDNA; 3106 BP.
XX
AC ABK12892;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 cDNA sequence.
XX
KM Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KM cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KM inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KM cell proliferative disorder; developmental disorder; epilepsy;
KM Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KM reproductive disorder; endometriosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 203..2851
FT 0 /*tag= a
FT /*product= "Human protease PRTS-9"
PN MO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001MO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
(INDY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM;
PI Deleagean AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Malia NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Rankumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kellick DA;
XX
DR WPI; 2002-090437/12.
DR P-PDB; AAU74749.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 5; Page 166-167; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present nucleic acid sequence encodes
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 3106 BP; 928 A; 633 C; 704 G; 841 T; 0 other;

```





```

FT      /*tag= a
XX      /product= "Human DPP8"
PN      MO200119866-A1.
XX      22-MAR-2001.
XX      11-SEP-2000; 2000MO-AU01085.
XX      10-SEP-1999; 99AU-0002762.
XX      18-FEB-2000; 2000AU-0005709.
XX      (UNSY ) UNIV SYDNEY.
XX      Abbott CA, Gorell MD;
XX      MPI; 2001-281520/29.
XX      P-PSDB; AAA47187.
XX      New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
XX      PT substrates, identifying inhibitors of DPP8 catalytic activity which
XX      PT have therapeutic uses, and for detecting activated T cells
XX      PS Claim 16; Fig 2; 78pp; English.
XX      This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8
XX      CC has substrate specificity for H-Ala-Pro-PNA, H-Gly-Pro-PNA and
XX      CC H-Arg-Pro-PNA. Therefore, it is a prolyl oligopeptidase and a
XX      CC dipeptidyl peptidase, because it is capable of hydrolyzing the
XX      CC peptide bond C-terminal to Pro in each of these compounds. DPP8
XX      CC is homologous with human DPPIV. DPP8 is useful for cleaving a
XX      CC substrate, and for detecting an activated T cell which involves
XX      CC measuring the level of DPP8 gene expression in a T cell. The level
XX      CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
XX      CC in the cell. It is also useful for identifying a molecule capable
XX      CC of inhibiting the cleavage of the substrate by DPP8. Molecules
XX      CC identified as inhibiting DPP8 catalytic activity may be useful for
XX      CC treating diabetes, growth hormone deficiency, lowering glucose levels
XX      CC in non-insulin dependent diabetes mellitus and other disorders
XX      CC involving glucose intolerance, enhancing mucosal regeneration and
XX      CC as immunosuppressants.
XX      SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Score: 7,45e-248 Length: 3120
Percent Similarity: 2422.00 Matches: 465
Best Local Similarity: 82.308 Conservative: 0
Query Match: 97.584 Mismatches: 0
DB: 22 Indels: 100 Gaps: 1

US-10-070-464-5 (1-465) x AAC85694 (1-3120)
QY      1 ThrGlyThrAlaAsnProLySValThrPheLySMeTSerGluIleMetIleAspAlaGlu 20
DB      1165 ACAGGTACGCAAAATCCTAAAGTCACTTTTAAGATGTCAGAAATATATGTTGTTGTA 1224
QY      21 GlyAglIleIleAspValIleAspLySgIuLeuIleGlnProPheGluIleuPheGlu 40
DB      1225 GGAAGATCATAGATGTCTATAGATAAGAACTAAATTCACCTTTGAGATTCATTTGAA 1284
QY      41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLySValTyrAlaIlePheSer 60
DB      1285 GGAGTTGATATATATGTCGAGACTGATGATGACTCTCTGAGGAGAAATATGCTGTGTCATC 1344
QY      61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
DB      1345 CTACTAGATCGCTCCACAGCTCGCTACAGATAGTGTGATCTCACCTGAATTAATTATTC 1404
QY      81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
DB      1405 CCAAGTAAGATGATGATTATGAGAAAGCAGAGACTCATTTGAGTCACTGCTGATTCCTG 1464

```

```

QY      101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120
DB      1465 ACGCCACTAATTAATCTATAGAGAAACACAGACATCTGATTAATATTCATGACATCTTT 1524
QY      121 HisValPheProGlnSerHisIleGluGluIleGluPheIlePheAlaSerGluCySlyS 140
DB      1525 CATGTTTCCCAAGAGTCAGAGAGAGAAATTGACATTAATTTTCCCTCTGATGACAA 1584
QY      141 ThrGlyPheArgHisIleuTyrIleValIleThrSerIleLeuLySgIuSerLySValArg 160
DB      1585 ACAGGTTTCGTCATTTATACAAATTAACATCATTTTAAAGGAAAGCAATTAACAGCA 1644
QY      161 SerSerGlyGlyLeuProAlaProSerAspPheLySValProIleLySgIuGluIleAla 180
DB      1645 TCCAGTGTGGGCTGCTCTCCAAAGTATTCAGATGTCATCAAGAGGATGATGCA 1704
QY      181 IleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValaPheGlu 200
DB      1705 ATTACCAAGTGGTGAATGGGAAGTTCTGGCCGCGCATGATCTTAATCCAAAGTTGATGAA 1764
QY      201 ValArgArgLeuValTyrPheGluGlyThrLySAspSerProLeuGluHisIleuTyr 220
DB      1765 GTCAGAGGCTGGTATATTTGAGGACCAAAAGACTCCCTTTAGAGCATCACTGTAC 1824
QY      221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240
DB      1825 GTAGTGTAGTATGATTAATCTGAGAGGTGACAGGCTACATGACCGTACATCTACAT 1884
QY      241 SerCySValSerGlnHisCySAspPhePheIleSerLySValSerAsnGlnLySAsn 260
DB      1885 TCTGTGTGATCAGTCAGACACTGTACTTTTATAGTATGATTAACCAAGAAAT 1944
QY      261 ProHisCySValSerIleuTyrLySLeuSerSerProGluAspAspProThrCySlyThr 280
DB      1945 CCACACTGTGTCTCTTTACCAAGCATCAAGTCCGAGAGATGACCAACTTCAAAAC 2004
QY      281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300
DB      2005 AAGAAATTTGGCCACCAATTTGATTCAGAGGCTCTCTTCGACATATACCTCTCCA 2064
QY      301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLySPro 320
DB      2065 GAAATTTCTCTTTGAAAGTACTGATTTACTTATATGAGATGCTCTCAAGCCT 2124
QY      321 HisAspLeuGlnProGlyLySlySValTyrProThrValLeuPheIleTyrGlyProGln 340
DB      2125 CATGATCTACAGCTCGAAAGAAATATCTACTGTGCTGTCAATATATGCTGCTCAG 2184
QY      340 ----- 340
DB      2185 GTGCAAGTGGTAATTAATCGTTTAAAGAGTCAAGATTTCCGCTTGAATACCTTACC 2244
QY      340 ----- 340
DB      2245 TCTCTAGTTATGTTGTTGTTAGTATGATACACACAGGGATCTGTCCAGGAGCTTAA 2304
QY      340 ----- 340
DB      2305 TTTGAAGCGCTTAAATATTAATATGATGGTCAATATAGAAATTGACAGATCAGGTGAAGA 2364
QY      340 ----- 340
DB      2365 CTCGAATATCTAGCTTCTCGATATGATTTCAATGACTTAGATCGTGTGGCATCCAGGC 2424
QY      340 ----- 340
DB      2425 TGTGCTATGAGAGATACCTCTCCCTGATGGCATTAATGACAGAGTCAGATATCTTCAGG 2484
QY      341 ValAlaIleAlaGlyValaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360
DB      2485 GTTGTATTTGTGGGGCCCAAGTCACTGTGTGATCTTCTATGATACAGGATACACGGA 2544

```

QY 361 ArgTyrMeGlyHisProAspGlnAsnGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2545 CGTATATGGGTCACTGACCGACAGAAATGAACAGGGCTATTACTAGGATCTGTGGCATG 2604  
 QY 381 GlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuHisGlyPheLeuAsp 400  
 Db 2605 CAAGCAGAAAAGTCCCTCGAACCAAAATCGTTTACTGCTTACATGCTTCCGAT 2664  
 QY 401 GlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2665 GAGAAATCTCCATTCTTGACATACAGATATATTACTGAGTTTAAAGTGAAGGCTGGAAG 2724  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGlnSerGly 440  
 Db 2725 CCATATGATATTACAGATCTATCTCCAGAGACACAGCATAAAGATTCCTGAATCCGGA 2784  
 QY 441 GlnHisTyrGlnLeuHisIleLeuHisTyrLeuGlnGlnLeuAsnGlySerArgIleAla 460  
 Db 2785 GAACATTATGACATGCTCTTTGCACTACCTCAAGAAAACCTTGGATCAGTATGCT 2844  
 QY 461 AlaLeuLysValIle 465  
 Db 2845 GCTCTAAAGTGATA 2859

## RESULT 11

AAD38956  
 ID AAD38956 standard; cDNA; 3120 BP.

AC AAD38956;

DT 23-SEP-2002 (first entry)

XX Human dipeptidyl peptidase 8 (DPP8) cDNA.

XX Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;  
 KM autoimmunity; human immuno deficiency virus; HIV infection; cytosolic;  
 KM graft rejection; antidiabetic; antiinflammatory; immunosuppressive;  
 KM antiviral; enzyme; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 214..2862

FT /tag= a

FT /product= "Human DPP8 protein"

XX WO200234900-A1.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-AU01388.

XX PR 27-OCT-2000; 2000AU-0001078.

XX (UNSY ) UNIV SYDNEY.

XX Abbott CA, Gorrell MD;

XX WPI; 2002-454646/48.

XX P-PSDB; AAE24170.

XX New dipeptidyl peptidase (DPP) peptides, useful for screening

XX PT inhibitors of DPP catalytic activity, which may be employed to treat

XX PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft

XX rejection and HIV infection -

XX Example; Fig 1; 91pp; English.

CC sequence is human DPP8 cDNA.

XX Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

XX SQ

Alignment Scores:

Pred. No.: 7.45e-248 Length: 3120

Score: 2422.00 Matches: 465

Percent Similarity: 82.30% Conservative: 0

Best Local Similarity: 82.30% Mismatches: 0

Query Match: 97.58% Indels: 100

DB: 24 Gaps: 1

US-10-070-464-5 (1-465) x AAD38956 (1-3120)

QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 1165 ACAGGTACAGCAAAATCTTAAGTACACTTTTAAGATGTCAGAAATTAAGATTGATGCTGAA 1224  
 QY 21 GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 1225 GSAAGCATCATAGTCATATAGATTAAGGAATTAATCACTTTTGAGATTCTATTGAA 1284  
 QY 41 GlyValGlnTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyrAlaTyrSerIle 60  
 Db 1285 GAGTTGAATATATGTCAGAGCTGATGATGATCTCGAGGAAATATGCTTGATCCATC 1344  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGlnLeuPheIle 80  
 Db 1345 CTACTGATGCTGCCAGACTCCGCTTACATATGTTGATGTCACCTGATTTATTC 1404  
 QY 81 ProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerValProAspSerVal 100  
 Db 1405 CCACTGAAGATGATGTTATGAGAAAGCAGACACTCTTGAAGCAGCTGATTCGTG 1464  
 QY 101 ThrProLeuIleIleTyrGlnGlnThrThrAspIleIleThrIleAsnIleHisAspIlePhe 120  
 Db 1465 ACGCCTAATATATCTATGAAAGAACACAGACTCTGATTAATATCTCATGACATCTTT 1524  
 QY 121 HisValPheProGlnSerHisGlnGlnGlnIleGlnPheIlePheAlaSerGlnCysLys 140  
 Db 1525 CATGTTTTTCCCAAGTCCAGAAAGGAATGAGTTATTTTGGCTCTGATGCAAA 1584  
 QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArg 160  
 Db 1585 ACAGTTTCCGTCATTTATCAAAATTAATCAATTTTAAAGAAACCAATTAATAA 1644  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGlnGluIleAla 180  
 Db 1645 TCCAGTGGTGGCTGCTCCAGAGTTCAGATTCAGTCTTATCAAAAGAGAGATAGCA 1704  
 QY 181 IleThrSerGlyGlnTyrGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
 Db 1705 ATTACCAAGTGTGATGGAAGTTCTTGGCGGATGATCTAATATCCAAAGTTGATGAA 1764  
 QY 201 ValArgArgLeuValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyr 220  
 Db 1765 GTCAGAAAGCTGGATATTTTGAAGGACCAAAAGACTCCCTTTAGAGCATCACTGTAC 1824  
 QY 221 ValValSerTyrValAsnProGlyGlnValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1825 GTAGTCAGTTACGTAATCTTGAAGAGTACAAAGGCTGACTGACCTGCTACTCAT 1884  
 QY 241 SerCysValIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260  
 Db 1885 TCTTGCTGATCAGTACGACTGTGACTTCTTATTAAGTAAATAGTAACTCAGAGAAAT 1944  
 QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGlnAspAspProThrCysLysThr 280  
 Db 1945 CCACACTGTGTGCTCCCTTACAAGCTATCAAGTCTGAAGATATCAACCACTCTCAAAACA 2004  
 QY 281 LysGlnPheThrPheAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2005 AAGGAATTTTGGGCCACATTTTGGATTACGACAGGTCCTTCTGACTATATCTCTCCA 2064

QY 301 Glu11ePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 DB 2065 GAATTTCTCTTTGAAAGACTACTGATTTACATTTGATGAGATGCTTACAAAGCTT 2124  
 QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
 DB 2125 CATGATCTACAGCTCGAAAGAAATATCTCTGCTGCTTATATATGATGCTGCTCAG 2184  
 QY 340 ----- 340  
 DB 2185 GTGCAATTGGTGAATTAATCGGTTTAAAGAGTCAAGTATTTCCGCTTGAATACCTAGCC 2244  
 QY 340 ----- 340  
 DB 2245 TCTCTAGTTATGTTGTTAGTATGATACAAAGAGGATCTCTTCAAGAGGCTTAA 2304  
 QY 340 ----- 340  
 DB 2305 TTTCAGAGCGCTTTAATATATAATGAGTCAATAGAAATTGACGATCAGTGGAAAGA 2364  
 QY 340 ----- 340  
 DB 2365 CTCCAATATATCTAGCTTCTCGATATGATTTCACTGACTTATGATCGTGGGATCCAGCGC 2424  
 QY 340 ----- 340  
 DB 2425 TGCTCTTATGAGAGATACCTCTCCCTGATGSCATTAATGACAGAGTCAGATATCTTCAAG 2484  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 DB 2485 GTTGCTATGCTGGGGCCCGACGATCACTCTGGATCTTCTATGATACAGATACACGGAA 2544  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 DB 2545 CGTTATATGGGTCACTGACCAAGATGAACAGGGCTATTAATGAGATCTGTGGCCATG 2604  
 QY 381 GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 DB 2605 CAAGCAAGAAAGTTCCCTCTGAAACCAATCGTTACTGCTTACATGGTTTCTTGAT 2664  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 DB 2665 GAGAAATGTCATTTTGGACATACACAGATATTAATGAGTTTATGAGAGGGCTGGAAG 2724  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGlnAlaArgHisSerIleArgValProGlnSerGly 440  
 DB 2725 CCAATGATTTACAGATCTATCTCAGAGAGACACAGCATTAAGAGTTCTGAAATCGGA 2784  
 QY 441 GluHisTyrGlyLeuHisLeuLeuHisTyrLeuGlnGlnAlaSerArgIleAla 460  
 DB 2785 GAACATTTATGACCTGATCTTTTGGACATCTTCAAGAAACCTTGGATCACGTTATGCT 2844  
 QY 461 AlaLeuLysValIle 465  
 DB 2845 GCTCTAAAGTGATA 2859  
 RESULT 12  
 AAH99934  
 ID AAH99934 standard; cDNA, 3143 BP.  
 AC AAH99934;  
 XX  
 XX 12-APR-2002 (first entry)  
 DT  
 XX  
 DE cDNA encoding 21953 human prollyl oligopeptidase.  
 XX  
 XX 21953 prollyl oligopeptidase; human; prollyl; endopeptidase;  
 KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;  
 KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;  
 KW antidiabetic; antiarthritic; antiaesthetic; antiinflammatory;  
 KW diabetes mellitus; arthritis; multiple sclerosis; asthma;  
 KW Grave's disease; neuronal disorder; demyelinating disease; ss.

XX OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 0 229..2877  
 FT /tag= a  
 FT /product= "21953 prollyl oligopeptidase"  
 FT /note= "This region is specifically claimed in  
 claim 2"  
 XX  
 XX MO200179473-A2.  
 XX  
 XX  
 XX 25-OCT-2001.  
 XX  
 XX 11-APR-2001; 2001WO-US40483.  
 XX  
 XX 18-APR-2000; 2000US-197508P.  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX  
 XX Meyers RA, Williamson M;  
 XX  
 XX WPI; 2002-034353/04.  
 DR P-PSDB; AAG78415.  
 XX  
 XX  
 PT New polypeptides 21953, member of human prollyl oligopeptidase family,  
 PT useful as diagnostic targets and therapeutic agents for controlling  
 PT cancer, lymphoma and leukemia  
 PS Claim 7; Page 100-102; 121pp; English.  
 XX  
 CC This invention relates to an isolated 21953 human prollyl  
 CC oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,  
 CC neoprotective, antihypertoid, dermatological, antipsoriatic,  
 CC antiaesthetic, ophthalmological, antiinflammatory, nootropic,  
 CC antiparkinsonian, anticoagulant, gynaecological, vasotropic,  
 CC antianginal, cardiant, antiatherosclerotic, anorectic and  
 CC metabolic in its action. Uses include gene therapy, expression or  
 CC activity of 21953 protein modulator, it is useful for identifying a  
 CC compound which binds to it and can be used in preventing, treating  
 CC or detecting a cellular proliferative or differentiative disorder.  
 CC The 21953 molecules can act as novel diagnostic targets and therapeutic  
 CC agents for controlling disorders associated with the aberrant activity  
 CC or degradation of peptide hormones e.g., disorders associated with cell  
 CC differentiation and proliferation such as cancer, immune function,  
 CC reproductive, neurological and cardiovascular function. The 21953  
 CC molecules are thus useful for treating and preventing cellular  
 CC proliferative and differentiative disorders, haematopoietic neoplastic  
 CC disorders, immune disorders such as autoimmune diseases, diabetes  
 CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease, and  
 CC neuronal disorders, demyelinating diseases, vascular disorders and  
 CC metabolism or pain disorders. This sequence represents the cDNA  
 CC encoding sequence of 21953 human prollyl oligopeptidase.  
 XX  
 SO Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7,54e-248 Length: 3143  
 Score: 2422.00 Matches: 465  
 Percent Similarity: 82.30% Conservative: 0  
 Best Local Similarity: 82.30% Mismatches: 0  
 Query Match: 97.58% Indels: 100  
 DB: 24 Gaps: 1  
 US-10-070-464-5 (1-465) x AAH99934 (1-3143)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 DB 1180 ACGGATACCAATCTTAAGATCTTTAAGATTCGAAATTAATGATGATGCTGAA 1239  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 DB 1240 GGAAGATCATGATCTCATAGTAAAGAACTAATTCACCTTTGAGATTCTATTGAA 1299

QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyIleTyrAlaATPSeTle 60  
 Db 1300 GGAGTGAATAATATTCGACAGCGTGATGAGATCTCTCAGGAAATATGCTGGCTCATC 1359  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1360 CTACTAGATGCTCCAGACTCGCTCGCATAGTGTGATCTTCACCTGAAATATTTATTC 1419  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 Db 1420 CCACTAGAACATGATGTTATGAAAAGCAGAGACTCATGAGTCAGCGCTGATTCGTG 1479  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 Db 1480 ACCCCACTAATTAATCTATGAAAGAAACAACAGACTCGGATTAATAATCCATGACATCTTT 1539  
 QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAspSerGluCysLeu 140  
 Db 1540 CATGTTTTTCCCAAGATCAGAAAGAGAAATGAGTTATTTTTTGGCTCTGAATGCATA 1599  
 QY 141 ThrGlyPheAspArgHisLeuTyrIleThrSerIleLeuIleGluSerIleTyrIleValArg 160  
 Db 1600 ACAGATTTCCGTCATTAATACAAAATTACATCTATTTTAAAGAAACCAATATTAACGA 1659  
 QY 161 SerSerGlyGlyLeuProAlaProSerAspPheCysProIleIleGluGluIleAla 180  
 Db 1660 TCCAGTGGTGGCTGCTGCTCCAAAGATTTCAAGTCTTCAAAAGAGAGATGACA 1719  
 QY 181 IleThrSerGlyGluTyrGluValLeuGlnValArgHisGlySerAsnIleGlnValAspGlu 200  
 Db 1720 ATTACCACTGATGATGGAATGGAAGTTCTTGCGCGCATGATCTAATATCCAAAGTTGATGAA 1779  
 QY 201 ValArgArgLeuValTyrPheGluGlyThrIleAspSerProLeuGlnHisIleTyr 220  
 Db 1780 GTCAGAAAGCGCTGGATATTTTGAAGGCACCAAGACTCCCTTAGAGCATCACTGTAC 1839  
 QY 221 ValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1840 GTAGTCAGTACGAAATCCGAGAGAGATGACAAAGGCTGACCTGCGCTACTCATCAT 1899  
 QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnIleValAsn 260  
 Db 1900 TCTTGCTGCATCACTGACGACTGACTCTTTAATAAGTATAGTATACCAAGAAAT 1959  
 QY 261 ProHisCysValSerLeuTyrIleLeuSerSerProGluAspAspProThrCysValSerThr 280  
 Db 1960 CCACACTGTGTCTCTTTACAAAGCTATCAAGTCTGAMAGATGACCAACTTGCAAAACA 2019  
 QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2020 AAGGAATTTGGCCACCATTTGGATTCAGACAGCTCTCTCTGACTATACTCTCCA 2079  
 QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIlePro 320  
 Db 2080 GAAATTTTCTTTTGAAGTACTACTGATTTACATGATGGAGCTCTACAAAGCT 2139  
 QY 321 HisAspLeuGlnProGlyIleValSerTyrProThrValLeuPheIleTyrGlyIleProGln 340  
 Db 2140 CATGATCTACAGCTGGAAGAAATATCTTCTGCTGCTCATATATGTTGTTGCTCCAG 2199  
 QY 340 ----- 340  
 Db 2200 GTGCACTGGTGAATAATCGTTTAAAGAGTCAAGTATTTCCGCTTGAATCCCTAAGCC 2259  
 QY 340 ----- 340  
 Db 2260 TCTCTAGTTATGTGTTAGTATAGTACAAACAGGGATCTGTCAACGAGGGCTTAA 2319  
 QY 340 ----- 340  
 Db 2320 TTGTAAAGCGCCTTTAAATATTAATGCGTCAAAATAGAAATGACATCAGTGAAGGA 2379

QY 340 ----- 340  
 Db 2380 CTCCAATATCTACCTTCTCGATATGATTCATTGACTAGATCGTGGCATCCACGCG 2439  
 QY 340 ----- 340  
 Db 2440 TGGTCTTAGGAGATACCTCTCCCTGATGCGATTAATGACAGTCAGATATCTTCAG 2499  
 QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
 Db 2500 GTTGCTATTGCTGGGCGCCAGTCACTCTGTGATCTTCTATGATACAGGATACAGGAA 2559  
 QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2560 CGTTATATGGTACCCCTGACCAAGATGAACAGAGGCTATTACTTAGATTCGTGGCATG 2619  
 QY 381 GluAlaGluIlePheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 Db 2620 GAGGATGTCATTTTGCACATACCAATATCTAGTGTCTTACATGAGGCTGGAAG 2679  
 QY 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLeu 420  
 Db 2680 GAGATGTCATTTTGCACATACCAATATCTAGTGTCTTACATGAGGCTGGAAG 2739  
 QY 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
 Db 2740 CCATATGATTTTACAGATCTATCTCCAGAGAGACACAGCTTAAGATTCCTGAAATGCGGA 2799  
 QY 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 Db 2800 GAACATTAATGACATGCAATCTTTTGCATCACTACCTTCAAGAAAACCTTGATACGATTTGCT 2859  
 QY 461 AlaLeuIleValIle 465  
 Db 2860 GCTCTAAAAGTGATA 2874  
 QY 461 AlaLeuIleValIle 465  
 Db 2860 GCTCTAAAAGTGATA 2874  
 RESULT 13  
 ABK83331  
 ID ABK83331 standard; cDNA; 4676 BP.  
 XX  
 AC ABK83331;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE cDNA encoding human DPP-1 splice variant #7.  
 XX  
 KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;  
 KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;  
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;  
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;  
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;  
 KM dyskinetia; reproductive disorder; inflammatory disorder;  
 KM metabolic disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200231134-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 12-OCT-2001; 2001MO-US31874.  
 XX  
 PR 12-OCT-2000; 2000US-240117P.  
 XX  
 PA (FERR) FERRING BV.  
 XX  
 PI QJ S, Akinsanya KO, Riviere PJ, Junien J;  
 XX  
 DR WPI; 2002-444178/47.  
 DR P-PSDB; ABG61600.  
 XX  
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding  
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and

PT viral infections, cancers, allergies, neurological disorders, or pain  
XX  
PS Disclosure; Page 72-73; 113pp; English.  
XX  
CC The present invention relates to the isolation of novel human serine  
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related  
CC proteins (DPP). The dipeptidyl peptidase IV-related proteins (DPP)  
CC and nucleic acids encoding them are useful for treating infections  
CC such as fungal, bacterial, protozoan and viral infections, particularly  
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),  
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,  
CC bulimia, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,  
CC psychotic and neurological disorders (e.g. anxiety, dementia, or  
CC schizophrenia), and dyskinesias. These may also be used in discovering  
CC therapeutic agents for the treatment of reproductive, inflammatory and  
CC metabolic disorders. ABK83322-ABK83343 encode human DPP proteins.  
XX

SO Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;

## Alignment Scores:

Pred. No.:	3,256-247	Length:	4676
Score:	2418.50	Matches:	463
Percent Similarity:	90.10%	Conservative:	1
Best Local Similarity:	89.90%	Mismatches:	1
Query Match:	97.44%	Indels:	51
DB:	24	Gaps:	1

US-10-070-464-5 (1-465) X ABK83331 (1-4676)

QY 1 ThrGlyThrAlaenProLysValThrPhelysmetserGluileMetileAspAlaGlu 20  
DB 1165 ACAGGTACAGCAATTCCTAAAGTCACCTTTAAGATGTCAGAAATAATGATGATGCGAA 1224  
QY 21 GlyArgileileaspeValileasplysGluileuileglnProphelGluileLeuPheGlu 40  
DB 1225 GGAAGATCATAGATGTCATAGATGATGAACTAATTCACCTTTGAGATTCATTTGAA 1284  
QY 41 GlyValGluTyrileAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerile 60  
DB 1285 GGAAGTAAATATATGTCAGAGCTGATGATGACCTCTGAGGAAATATGCTGCTCATC 1344  
QY 61 LeuLeuAspArgserGlnThrArgleuGlnileValleuileserProGluLeuPheile 80  
DB 1345 CTACTAAGTCGCTCCAGACTCGCTACAGATAGTGTGATCTCACTGAAATATTTATC 1404  
QY 81 ProValGluAspAspValMetGluArgGlnArgleuileGluSerValProAspSerVal 100  
DB 1405 CCAAGTAGAAGATGATGTTATGAAAGGAGAGACTCATTTGATGATGAGTCTGATTCGTG 1464  
QY 101 ThrProleuileileTyrGluGluThrThrAspilleTyrPheleuileAspillePhe 120  
DB 1465 ACCGCACTAATTAATCTATGAAAGAAACAACACATCTGATTAATATCCATGACATCTTT 1524  
QY 121 HisValPheProGlnSerHisGluGluGlnileGluPheilePheAlaSerGluCysVal 140  
DB 1525 CATGTTTTTCCCAAAGTCACGAAAGGAAATTAAGTTTATTTTGGCTCTGAAATGCAAA 1584  
QY 141 ThrGlyPheArgHisLeuTyrLysileThrSerileLeuLysGluSerLysTyrLysArg 160  
DB 1585 ACAAGTTTCCGTCATTAACAATAATCATCTATTTTAAAGGAAACAATAATTAACGA 1644  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProilleLysGluGluileAla 180  
DB 1645 TCCAGTGTGGGCTGCTGCTCCAAAGATTTCAAGGTCTCTATCAAAAGGAGATGACA 1704  
QY 181 IleThrSerGlyGluTyrPGluValleuGlyArgHisGlySerAsnileGlnValAspGlu 200  
DB 1705 ATTAACGATGTGAATGGGAAGTTCTTGCGCGGATGATGATTAATTCGAAAGTTGAGAA 1764  
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeuTyr 220

DB 1765 GTCAAGAGGCTGGATATTTTGAAGGACCAAAAGCTCCCTTTAGAGCATCAGCTGAC 1824  
QY 221 ValValSerTyrValaenProGlyGluValThrArgleuThrAspArgGlyTyrSerHis 240  
DB 1825 GTACTCAGTAAATGAAATCCCTGGAGAGTGCAGAGGCTGACCTGCGCTACTACAT 1884  
QY 241 SerGlyCysIleSerGlnHisCysAspPhePheileSerLysTyrSerAsnileLysAsn 260  
DB 1885 TCTTGCTGACATGACATGACATGACATCTTTTAAAGTAAATGATTAACCAAGAAAT 1944  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1945 CCAACATGTCGTCTCTTTCAAGCTATCAAGTCTGAAGTGAACCAACTTGCAGAAACA 2004  
QY 281 LysGluPheThrAlaThrileLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
DB 2005 AAGCAATTTTGGGACACCATTTGGATTCAGACAGTCTCTTCTGACTATATCTCTCA 2064  
QY 301 GluilePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 2065 GAAATTTCTCTTTGAAAGTACTGAGATTAATGATTAATGAGATGCTTACAAAGCT 2124  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValleuPheileTyrGlyGlyPro--- 339  
DB 2125 CATGATCTACAGCTGAGAAAGAAATATCTTACTGCTGTTCAATATGTCGTGCG- GCT 2183  
QY 339 ----- 339  
DB 2184 CAATAGAAATTCAGATGAGTGAAGAGACTCCAAATATGATCTCTCGATATGATTTTC 2243  
QY 339 ----- 339  
DB 2244 ATTGACTTAAGATGTCGTGGCATTCACGCTGCTCTATGAGAGATACCTCTCCGATG 2303  
QY 340 -----GlnValAlaileAlaGlyAlaProValThrleu 350  
DB 2304 GCATTAAATGACAGATGATATCTTCAAGGCTTCTATTCCTGGGCTCCAGTCACTCTG 2363  
QY 351 TrrPhePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisAspAspGlnAsnGlu 370  
DB 2364 TGGATCTTCTATGATACAGATACACGAAAGTATATGGGTGACCTGACCAAGATGAA 2423  
QY 371 GlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProAspSerGluProAsn 390  
DB 2424 CAGGCTAATTAATTAAGATGTCGTGGCCATCAAGCAAGAAAGTCCCTCTGAAACCAAT 2483  
QY 391 ArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheAlaHisThrSerile 410  
DB 2484 CGTTTACTGCTCTTACATGCTTCTGATGAGATGATGATGATGATGATGATGATGATG 2543  
QY 411 LeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnileTyrProGlnGlu 430  
DB 2544 TTAAGTATTTTAAAGAGGCTGAGAAACCATATGATTAACGATCTATCTCAGAGAG 2603  
QY 431 ArgHisSerileArgValProGluSerGlyGluHisTyrGluLeuHisLeuHisTyr 450  
DB 2604 AGACACAGCATTAAGATGCTTGAATCGGAGAAATATTAACATGATCTTTTGACATAC 2663  
QY 451 LeuGlnGluAsnLeuGlySerArgileAlaAlaLeuLysValile 465  
DB 2664 CTTCAAGAAACCTTGGATGACGATATGCTGCTCTAAGATGATTA 2708  
RESULT 14  
ABK83327 standard; cDNA; 4829 BP.  
ID ABK83327  
XX  
AC ABK83327;  
XX  
DT 12-AUG-2002 (first entry)  
XX  
DE cDNA encoding human DPP-1 splice variant #3.  
XX



KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPP; DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain; diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; heart failure; hypertension; urinary retention; osteoporosis; cancer; ulcer; allergy; cancer; psychotic disorder; neurological disorder; dyskinnesia; reproductive disorder; inflammatory disorder; metabolic disorder; gene; ss.

XX Homo sapiens.

OS WO200231134-A2.

PN 18-APR-2002.

PD 12-OCT-2001; 2001WO-US31874.

PF 12-OCT-2000; 2000US-240117P.

PR (FERR ) FERRING BV.

PA Qi S, Akinsanya KO, Riviere PJ, Junien J;

PI WPI: 2002-444178/47.

DR P-PSDB: ABG61596.

XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain

PT -

XX Disclosure; Page 65-66; 113pp; English.

XX The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related proteins (DPP). The dipeptidyl peptidase IV-related proteins (DPP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, CC psychotic and neurological disorders (e.g. anxiety, dementia, or CC schizophrenia), and dyskinasias. These may also be used in discovering therapeutic agents for the treatment of reproductive, inflammatory and CC metabolic disorders. ABK83322-ABK83343 encode human DPP proteins.

XX SO Sequence 4829 BP; 1466 A; 886 C; 1017 G; 1460 T; 0 other.

Alignment Scores:

Pred. No.:	1.97e-245	Length:	4829
Score:	2402.00	Matches:	465
Percent Similarity:	82.01%	Conservative:	0
Best Local Similarity:	82.01%	Mismatches:	0
Query Match:	96.78%	Indels:	102
DB:	24	Gaps:	1

US-10-070-464-5 (1-465) x ABK83327 (1-4829)

QY 1 ThrGlyThralaAsnProlyValThPhelysweSerGluIleMetIleAspAlaGlu 20

Db 1165 ACGAGTACGCAATCTTAAGTCATTTTAAGATGCAAGAAATTAAGATTGATCGTGA 1224

QY 21 GlyArgIleIleAspValIleAspLyseGluLeuIleGlnProheGluIleAspHeGlu 40

Db 1225 GGAAGGATCATAGATGTCATAGATTAAGAACTAATTCACCTTTTGATTCATATTGAA 1284

QY 41 GlyValGluIleValIleAlaAspAlaGlyTyrThrProGluGlyLeuTyrAlaAspSerIle 60

Db 1285 GGAGTTGAATATATGCGCAAGCTGAGATGACCTCGAGGAAATATATGCTTGCTCATC 1344

QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80

Db 1345 CTACTAGATGCTCCCAAGCTCCGCTACAGATAGTGTTCACCTGAATATTATTC 1404

QY 81 ProValGluAspAspValIleMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100

Db 1405 CCACTAAGATAGATGTTATGAAAGGCAAGACTCTTACGACAGCTCGATTCGTG 1464

QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120

Db 1466 ACCGCACTAATTTATCTATGAAGAAACAACAGACTCGATTAATATCCATGACATCTTT 1524

QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeu 140

Db 1525 CATGTTTTCCTCCCAAGTCAAGAAAGAAATGAGTTATTTTGTCTCGAATGCAAA 1564

QY 141 ThrGlyPheArgHisLeuTyrIleValIleThrSerIleLeuArgGluSerIleTyrLeuArg 160

Db 1585 ACAAGTTCCGCTCATTTATACAAATTAATCATTTTAAAGAAACCAATATTAACGA 1644

QY 161 SerSerGlyGlyLeuProAlaProSerAspPheIleCysProIleIleGluGluIleAla 180

Db 1645 TCCAGTGTGGCTGCTGCTCCAAAGTATTCAGTCTCATCAAAAGAGAGATAGCA 1704

QY 181 IleThrSerGlyGluTyrGluValIleGluIleArgHisGlySerAsnIleGlnValAspGlu 200

Db 1705 ATTAACAGTGTGAATGGAAGTTCTTGCGCGCATGATCTAATATCCAAAGTTGATGAA 1764

QY 201 ValArgArgLeuValTyrPheGluGlyThrLeuAspSerProLeuGlnHisIleLeuTyr 220

Db 1765 GTCAGAAAGCTGTATTTTGAAGGACCAAGACCTCCCTTAGACATCACTGTAC 1824

QY 221 ValIleSerTyrValAsnProGlyGluValIleThrArgLeuThrAspArgIleTyrSerHis 240

Db 1825 GTAGTCAGTACCTAATACCTGAGAGAGTACAAAGCTGACCTGCTCATCAT 1884

QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerArgGlnIleValAsn 260

Db 1885 TCTTGCTGATCTGACGACGCTGACCTGCTTTTATAGTATGATATACCAAGAAAT 1944

QY 261 ProHisCysValSerLeuTyrIleLeuAspSerIleGlyProLeuProAspIleTyrProPro 280

Db 1945 CCAACGTGTGTCCCTTACCAAGCTATCAAGCTGAAATGACCACTTGCAAAACA 2004

QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspIleTyrProPro 300

Db 2005 AAGAAATTTGGGCAACCATTTGGATTCAGCAGGCTCTCTCTGACTATACCTCCA 2064

QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIleValPro 320

Db 2065 GAAATTTCTCTTTGAAAGTACTGATTAATTAATGATGATGATGCTCTACAAAGCT 2124

QY 321 HisAspLeuGlnProGlyIleValIleValIleValIleValIleValIleValIleValIle 340

Db 2125 CATGATCTAAGCTGGAAGAAATATCCATGCTGTGTTCTATATATGATGATGCTCTCT 2184

QY 340 1n----- 340

Db 2185 AGGTGAGTTGTGAATTAATCGGTTAAAGAGTCAAGTATTCGCTGAATACCTGAG 2244

QY 340 ----- 340

Db 2245 CCTCTAGGTTATGTGTTGTATGATATGACAAACAGGGAGTCTGTACCGAGGCTTA 2304

QY 340 ----- 340

Db 2305 AATTGAAGCGCTTTAATATATAAATGAGTCAAAATGAAATGACATCAGGTGAAG 2364

QY 340 ----- 340

Db 2365 GACTCAATATCTAGCTTCTCATATGATTTATGACTTATGATGCTGTGGGATCAGCAG 2424

QY 340 ----- 340

Db 2425 GCTGATCTATGAGATATACCTCTCCCTGATGCAATTAATGACAGAGTCAATATCTTCA 2484



QY 341 --ValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrG 360  
 Db 2485 GGSTTGCTATTCGCGGGGCCCACTCACTCTGTGATCTTCTATGATACAGATACACGG 2544  
 QY 360 IuArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaM 380  
 Db 2545 AACGTTATATGGGTGACCCCTGACCAAGATGAACAGGGCTATTACTTGAATCTGTGGCCA 2604  
 QY 380 eGlnIleGlyIlePheProSerGlnProAsnArgLeuLeuLeuLeuHisGlyPheLeuA 400  
 Db 2605 TGCAGAGAGAAAGTTCCTCCCTGACCAATCGTTACTGCTTACATGATGTTCTCGG 2664  
 QY 400 spGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyL 420  
 Db 2665 ATGAGATGTCATCTTTGACATACCAATATTAATGATTTTATGAGGGCTGGAA 2724  
 QY 420 YSPProTyrAspLeuGlnIleTyrProGlnGlnArgHisSerIleArgValProGlnSerG 440  
 Db 2725 AGCCATATGATTTTCATCATCTATCTCAGAGAGACACAGCATAGAGATTCTGAATCGG 2784  
 QY 440 IyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnGluAsnLeuGlySerArgIleA 460  
 Db 2785 GAGAAACATTAATGAACCTCATCTTTTGCACTTCAAGAAACCTTGATCAGATATGG 2844  
 QY 460 IAlaLeuLysValIle 465  
 Db 2845 CTCCTCTAAAAGTCATA 2861  
 RESULT 15  
 AAD23843  
 ID AAD23843 standard; cDNA; 2510 BP.  
 XX  
 AC AAD23843;  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Human protease PRTS-2 cDNA.  
 XX  
 KM Human; protease; PRTS-2; tranquilliser; gene therapy; vaccine; allergy;  
 KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
 KM atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
 KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
 KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;  
 KM hypertension; neurological disorder; Parkinson's disease; drug screening;  
 KM cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
 KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
 KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
 KM developmental disorder; reproductive disorder; infertility; diarrhoea;  
 KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 616..2358  
 FT /\*tag= a  
 FT /product= "Human protease PRTS-2 protein"  
 FT  
 FT  
 PN MO200183775-A2.  
 XX  
 XX 08-NOV-2001.  
 PD  
 XX  
 PF 04-MAY-2001; 2001MO-US14651.  
 XX  
 PR 04-MAY-2000; 2000US-202082P.  
 PR 11-MAY-2000; 2000US-203566P.  
 PR 17-MAY-2000; 2000US-205803P.  
 PR 25-MAY-2000; 2000US-207477P.  
 PR 01-JUN-2000; 2000US-209402P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Deleageane AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L,

PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;  
 PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM;  
 PI Reddy R, Yue H, Tang YT;  
 DR WPI; 2002-034518/04.  
 DR P-PSDB; AAE14337.  
 XX  
 PT Novel human proteases and polymucleotides encoding the proteases,  
 PT useful for treating, diagnosing or preventing cell proliferative,  
 PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
 PT disorders -  
 XX  
 PS Claim 5; Page 139-140; 151pp; English.  
 XX  
 CC The invention relates to human proteases (PRTS-14) and its corresponding  
 CC cDNA molecule. Human PRTS and its nucleic acid molecule are useful for  
 CC the diagnosis, treatment and prevention of disorders associated with  
 CC increased or decreased expression of PRTS. Examples of such disorders  
 CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
 CC hepatitis, psoriasis and cancer); autoimmune/inflammatory disorders  
 CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,  
 CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
 CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
 CC viral, bacterial, fungal, parasitic, protozoal and helminthic  
 CC infections); cardiovascular disorders (myocardial infarction, ischemic  
 CC heart disease and hypertension); neurological disorders (epilepsy,  
 CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,  
 CC Parkinson's disease, stroke, mental disorders including mood, anxiety  
 CC and seasonal affective disorder and prion diseases); gastrointestinal  
 CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
 CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
 CC alopecia, scabies, insect bites and urticaria); reproductive disorder  
 CC (infertility, disruption of estrous and menstrual cycle and  
 CC gynaecomastia); and developmental disorders (renal tubular acidosis,  
 CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
 CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic  
 CC fragments are useful for screening libraries of compounds in conventional drug  
 CC screening assays. The present sequence is human protease PRTS-2 cDNA.  
 XX  
 SO Sequence 2510 BP; 777 A; 494 C; 527 G; 712 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4.84e-229 Length: 2510  
 Score: 2244.50 Matches: 439  
 Percent Similarity: 84.03% Mismatches: 3  
 Best Local Similarity: 83.46% Indels: 5  
 Query Match: 90.43% Gaps: 6  
 DB: 24  
 US-10-070-464-5 (1-465) x AAD23843 (1-2510)  
 QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 822 ACAGGTACAGCAATCTTAAGTCACCTTTAAAGATGCACAAATAAGATTGATGCTGAA 891  
 QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
 Db 892 GGAAGATCATATGATGATCAATGAACTAATTAACCTTTTGAGATTCTATTGAA 951  
 QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyrAlaTyrSerIle 60  
 Db 952 GGAAGTGAATATATTCGACAGCTGAGTGAACCTCGAGGAAATAATGCTTGTCATC 1011  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleLeuSerProGluLeuPheIle 80  
 Db 1012 CTACTGATTCGCTCCAGACTCCCTACAGATGATGTTGATCCACTGGAATATTATTC 1071  
 QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGlnSerValProAspSerVal 100  
 Db 1072 CCACTGAAATATATGTTATGCAAGCAAGCATCTTTAGTCAGGCTGATTCGTG 1131  
 QY 101 ThrProLeuIleIleTyrGlnGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120

Db 1132 ACGCCACTAATTATCTATGAAGAAACAAGACATCGGATTAATATCCATGACATCTTT 1191  
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeu 140  
Db 1192 CATGTTTTTCCCAAGTACAGAAAGAAATGAGTTATTTTGGCTTGAATGCAA 1251  
Qy 141 ThrGlyPheArgHisLeuTyrIleThrSerIleLeuGluSerLeuTyrIleArg 160  
Db 1252 ACAGGTTCCGTCATTTATACAAAATATCATCTATTTTAAAGAAACAAATATATAACA 1311  
Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLeuGluGluIleAla 180  
Db 1312 TCAGTGGTGGGCTGCTGCTCAACT----- 1338  
Qy 181 IleThrSerGlyGluTyrGluVal-----LeuGly----- 190  
Db 1339 GTCACT-----TGATGATGATCATTCATGAGATCTCTAGGAACCTCCGTATG 1389  
Qy 191 -----ArgHisGlySerAsnIleGluValArgArgLeuValTyrPheGlu 208  
Db 1390 TGTGTGACACATATAGTTAGATCCAGTGTGATGAAAGTCAGAGGCTGTATATTGAA 1449  
Qy 209 GlyThrIysAspSerProLeuGluHisIleuTyrValIleSerTyrValAsnProGly 228  
Db 1450 GGACACCAAGACTCCCTTTAGAGCATCACTGTAGTAGTACGTAATCTGGA 1509  
Qy 229 GluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGlnHisCys 248  
Db 1510 GAGGTGACAAAGCTGACGACGCTGCTACTCATCTTGTGTCATCAGTCAGCATGT 1569  
Qy 249 AspPhePheIleSerLeuTyrSerAsnGlnIysAsnProHisCysValSerLeuTyrIys 268  
Db 1570 GACTTCTTTATAGTAAGTATAGTAAACAGAAATCCACACTGTGTCTCCCTTACAAAG 1629  
Qy 269 LeuSerSerProGluAspAspProThrCysIysThrIysGluPheThrAlaThrIleLeu 288  
Db 1630 CTATCAAGCTCTGAGATGAGCCCACTGCAAAACAAAGAAATTTGGCCACCAATTTG 1689  
Qy 289 AspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSerThr 308  
Db 1690 GATTGACGAGGCTCTTCTGACATATCTCTCCAGAAATTTCTTTGAAAGTACT 1749  
Qy 309 ThrGlyPheThrLeuTyrGlyMetLeuTyrIysProHisAspLeuGlnProGlyIysIys 328  
Db 1750 ACTGATTTACATGTATGAGGATGCTCAAGCCTCATGATCTACAGCTGGAAGAAA 1809  
Qy 329 TyrProThrValIleuPheIleTyrGlyProGln----- 340  
Db 1810 TATCTACTGTGCTGTTATATATGATGCTCTCAAGTGCAGTGTGTAATATGCGTTT 1869  
Qy 340 ----- 340  
Db 1870 AAAGGAGTCAAGTATTTCCGCTTGAATACCTAGCCTCTAGGTTATGTGGTTAGTG 1929  
Qy 340 ----- 340  
Db 1930 ATAGACAAACAGGGGATCTGTACCGAGGCTTAAATTGAAGGCGCTTAAATATAAA 1989  
Qy 341 ---ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 359  
Db 1990 ATGCTGCTATGCTCGGGGCCCACTCACTGTGATCTTCTATGATACAGATACAG 2049  
Qy 360 GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379  
Db 2050 GAACGTTATATGGGTACCTGACCAAGATGAACAGGGCTATTACTTAGATCTGTGGCC 2109  
Qy 380 MetGlnAlaGluIysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399  
Db 2110 ATGCAAGCAAAAGTCCCTCTGACCAAAATCGTTACGCTTTACATGGATTCTTG 2169  
Qy 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419  
Db 2170 GATGGAATGTCAATTGGACATACAGATATATTACTGAGTTTATAGTGAAGGCTGGA 2229

Qy 420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439  
Db 2230 AAGCATATGATTTA-----CAGGAGAGACACAGCATTAAGAGTTCTGGAATCG 2277  
Qy 440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIle 459  
Db 2278 GGAGAACATATATGAACCTGCACTTTTGCACCTTCAAGAAAACCTTGATCACGTAAT 2337  
Qy 460 AlaAlaLeuLysValIle 465  
Db 2338 CTTGCTCTTAAGATGATA 2355

Search completed: October 15, 2003, 17:57:56  
Job time : 385.661 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: October 15, 2003, 17:29:02 | Search time 72.8508 Seconds  
(without alignments)  
2817.309 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTFKMSIWDAA.....HLHYDQENIGSRIALKVI 465

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10070464/runat.15102003.113554.24845/app.query.fasta\_1.2652  
-DB=Issued Patents NA -QFMT=fastag -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=400 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptio -NORM=none -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10070464 @CIGN 1.1 146 @runat.15102003.113554.24845 -NCP1=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfillseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	510.5	20.6	823	4	US-09-280-116-171
2	309.5	12.5	5496	4	US-09-462-284-1
3	293	11.8	612	4	US-09-392-184-31
4	238	9.6	3407	4	US-10-002-593-5
5	234	9.4	2924	5	PCT-US93-07923-1
6	214	8.6	2815	1	US-08-230-491A-1
7	214	8.6	2815	2	US-08-619-280A-1
8	193	7.8	4982	3	US-08-940-391-1
9	193	7.8	4982	3	US-08-699-103B-1
10	193	7.8	4982	4	US-09-229-059-1
11	193	7.8	4982	4	US-09-628-133-1
12	167	6.7	502	4	US-09-280-116-172

13	146.5	5.9	815	4	US-09-221-017B-962	Sequence 962, App
14	139	5.6	1896	3	US-09-016-080-2	Sequence 2, App1
15	132.5	5.3	543	4	US-09-221-017B-253	Sequence 253, App
16	124.5	5.0	657	4	US-09-221-017B-646	Sequence 646, App
17	110.5	4.5	3085	4	US-09-221-017B-1010	Sequence 1010, App
18	103.5	4.2	3942	3	US-09-162-484-19	Sequence 19, App1
19	103	4.1	904	4	US-09-221-017B-58	Sequence 58, App1
20	100.5	4.0	5798	2	US-08-658-665-178	Sequence 178, App
21	100.5	4.0	5798	3	US-08-796-101-39	Sequence 39, App
22	100	4.0	1674	1	US-08-365-981-2	Sequence 2, App1
23	96.5	3.9	2424	1	US-08-821-119-16	Sequence 16, App1
24	95	3.8	679	4	US-09-221-017B-13	Sequence 13, App1
25	95	3.8	1664976	4	US-08-916-421B-1	Sequence 1, App1
26	94.5	3.8	7215	4	US-09-134-001C-627	Sequence 627, App
27	93	3.7	1230025	4	US-09-198-452A-1	Sequence 1, App1
28	92.5	3.7	535	4	US-09-389-681-428	Sequence 428, App
29	92.5	3.7	535	4	US-09-620-405B-428	Sequence 428, App
30	92.5	3.7	535	4	US-09-433-826B-428	Sequence 428, App
31	92.5	3.7	535	4	US-09-604-287A-428	Sequence 428, App
32	92.5	3.7	2406	1	US-08-396-479B-7	Sequence 7, App1
33	92.5	3.7	2406	1	US-08-818-823-7	Sequence 7, App1
34	92.5	3.7	2647	1	US-08-396-479B-9	Sequence 9, App1
35	92.5	3.7	2647	1	US-08-818-823-9	Sequence 9, App1
36	92.5	3.7	3969	1	US-08-396-479B-11	Sequence 11, App1
37	92.5	3.7	3969	1	US-08-818-823-11	Sequence 11, App1
38	92.5	3.7	4010	5	PCT-US94-07297-40	Sequence 40, App1
39	92.5	3.7	19446	4	US-08-861-527-51	Sequence 51, App1
40	92	3.7	12295	4	US-09-221-017B-889	Sequence 889, App
41	90.5	3.6	1404	4	US-09-134-001C-615	Sequence 615, App
42	90.5	3.6	5883	1	US-08-404-445-1	Sequence 1, App1
43	90	3.6	1416	1	US-08-236-311-3	Sequence 3, App1
44	90	3.6	1416	3	US-08-457-918-3	Sequence 3, App1
45	90	3.6	1508	1	US-08-236-311-6	Sequence 6, App1

## ALIGNMENTS

RESULT 1  
US-09-280-116-171  
Sequence 171, Application US/09280116A  
Patent No. 6331427  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs  
FILE REFERENCE: 5800-24, 035800/176965  
CURRENT APPLICATION NUMBER: US/09/280,116A  
CURRENT FILING DATE: 1999-03-26  
NUMBER OF SEQ ID NOS: 268  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 171  
LENGTH: 823  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: prolyl oligopeptidases  
US-09-280-116-171

## Alignment Scores:

Pred. No.: 9.22e-51 Length: 823  
Score: 510.50 Matches: 115  
Percent Similarity: 85.71% Conservative: 5  
Best Local Similarity: 82.14% Mismatches: 5  
Query Match: 20.57% Indels: 15  
DB: 4 Gaps: 2  
US-10-070-464-5 (1-465) x US-09-280-116-171 (1-823)

QY 340 GINVAIAIAIAIAAGLYAIAProVAIAThrlaUTPILiePhetYrAspThrgLYTYrThr 359  
DB 151 AAGGTTGCTATTGCTGGGATCCAGACACTGTGTGATCTTCAATGATACAGATACAG 210  
QY 360 GUAAGTYrMeGLYHIsProAspGlnsngIUGnGLYTYrTYrleugLYserValaIA 379  
|||||



```
Db 3815 GGTCGTTTCACATTAGAGATCATACCGCCCTGTTTCTGACTGGCGTTTCTACGACTCA 3874
Qy 357 G1YTYrThGluAryrMetG1YH1sPProAspGluAsnGluGlnG1YTYrTyr 376
Db 3875 ATGTACACGGAGGCTACTAGAAAGACCCCTTCGACCATAGAGAGGCTAGAGACCAGC 3934
Qy 377 SerValAlaMetGlnAlaGluYrPheProSerGluProAsnArgLeuLeuHis 396
Db 3935 GCCGTC---CGAAGACTGACGGCTTCAAGACGTCGAGGCGGATTTCTTGATCCGAC 3991
Qy 397 GlyPheLeuAspGluAsnValHisPheAlaHisThSerIleLeuLeuSerPheLeuVal 416
Db 3992 GGAAACGGCGGACATACGTCATTCAGAACCTCGGCTCGCTGATCTCTGATG 4051
Qy 417 ArgAlaGlyLeuProTyraPheAsnGlnIleTyrProGlnGluArgHisSerIleArgVal 436
Db 4052 GCGGATGAC-----GTTCTTCCTGAGAGCTCCATTCGCAATGGCTTC 4093
Qy 437 ProGluSerGlyGluHisTyrgluLeuHis 446
Db 4094 ACAGACTCAGACGACGAAATCAGCTACCAT 4123

RESULT 3
US-09-392-184-31/C
; Sequence 31, Application US/09392184
; Patent No. 6395869
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392,184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FaSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; NAME/KEY: misc_feature
; LOCATION: (1)...(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31

Alignment Scores:
Pred. No.: 3,17e-25 Length: 612
Score: 293.00 Matches: 61
Percent Similarity: 66.35% Conservative: 8
Best Local Similarity: 58.65% Mismatches: 35
Query Match: 11.80% Indels: 0
Gaps: 0

US-10-070-464-5 (1-465) x US-09-392-184-31 (1-612)
Qy 340 GlnValAlaIleAlaGlyAlaProValIleLeuTyrIlePheTyraPheThrGlyTyrThr 359
Db 378 CAGGGTTCTATTTCTGGGGCCGACGTCATCTGTGATCTTCTATGAAACAGGAAACNGC 319
Qy 360 GluArgTyMetGlyHisPProAspGluAsnGluGlnG1YTYrTyrLeuGlySerValAla 379
Db 318 GGAAGTTAATGGGTACCCCTTNNCCAGNATGACAGGCGCTATTACTTGAATCTGTGGCC 259
Qy 380 MetGlnAlaGluYrPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
Db 258 ATTCACNACAGAAAGTTCCCTCTGNAACCAATNNGTNCCTCTTANATGGTTTCG 199
Qy 400 AspGluAsnValHisPheAlaHisThSerIleLeuLeuSerPheLeuValArgAlaGly 419
Db 198 GATTAGGATTTCCATTTTTCACANACCAGNANANTACTNAGGTTTTAGAGGCGCTGGG 139
```

```
Qy 420 LysProTyraPheAsnGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
Db 138 AAACCAATGATTTTNCAGAACCAACTTCNGAGAAAGNCCAAANAGGTTNCTGAATCG 79
Qy 440 GlyGluHisTy 443
Db 78 GGAGACAAATAT 67

RESULT 4
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Brown, Nancy J.
; TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERT)
; FILE REFERENCE: Acty Docket No. 6586198 1242/48/2
; CURRENT APPLICATION NUMBER: US/10/002,593
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/244,524
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 3407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.: 1.75e-17 Length: 3407
Score: 238.00 Matches: 120
Percent Similarity: 33.51% Conservative: 71
Best Local Similarity: 21.05% Mismatches: 193
Query Match: 9.59% Indels: 186
Gaps: 24

US-10-070-464-5 (1-465) x US-10-002-593-5 (1-3407)
Qy 2 GlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGly 21
Db 853 GGAGCTGTGATATCACTGTAAGTCTTTGTT-----GTAATACAGCTCTTCAGC 906
Qy 22 ArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGly 41
Db 907 TCAGTCACCAATGCACATTCATCAATCATCATGCTCTGCTTCTATGTGATA---GGG 963
Qy 42 ValGluTyrlleAlaArgAlaGlyTyrThrProGluGlyLysTyraIleAspSerIleLeu 61
Db 964 GATCACTACTTGTGTGATGTGATGCAATGGCA----- 993
Qy 62 LeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIlePro 81
Db 994 -----ACACAGAAAGAAATTTCTTGCAGTGC----- 1023
Qy 82 ValGluAspAspValMetGluArgGluArgLeuIleGluSerValProAspSerValThr 101
Db 1024 -----ACGAGGATTCAGAACTAT-----TCGGTCAG 1050
Qy 102 ProLeuIleIleTyrgluGluThrThrAspIleTyrIleAsnIleHisAspIlePheHis 121
Db 1051 GATATTGTGACTATGATATATCCAGTGAAGATG---AATGC----- 1092
Qy 122 ValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThr 141
Db 1093 TTAGTGGCAGCGGACACACTTGAATGACTACTGCTGGCTGGAGATTAGGCT 1152
Qy 142 GlyPheArgHis-----LeuTyrlleIleThSerIleLeuLys 154
Db 1153 TCAGAACTCATTTTAACTCTTGATGTGTAATAGCTTCAAGATCATCAGC-----AAT 1206
```

```

Qy 155 GluSerIyTyLysArgSerSeryGlyLeuProAlaProSerAspPheLysCysPro 174
    |||
    |||
    |||
Db 1207 GAAGAGAGTTACAGACACAT-----TCCTATTCCAATAGAT 1245
Qy 175 IleuGluGluIleAlaIleThrSeryGlyuTrpGluValLeuGlyArgHisGlySer 194
    |||
    |||
    |||
Db 1246 AAAAAGAGCTGCACATTTATTACAAAAGGACCTGGGAAGTCATCGG-----1293
Qy 195 AsnIleGluValAspGluValArgArgLeuValTyPheGluGlyThr---LysAspSer 213
    |||
    |||
    |||
Db 1294 ---ATAGAGCTTACACAGATGATTATCTATACCTAGTAAATGATATAAAGATG 1350
Qy 214 ProLeuGluHisIleuTyValValSeryTyValAsnProGlyGluValThrArgLeu 233
    |||
    |||
    |||
Db 1351 CCAGAGAGAGAAATCTTATTAAATCCAACTTATGACTTACAAAAGACATGCTCTC 1410
Qy 234 ThrAspArgGlyTySerHisSeryCysIleSer---GlnHisCysAspPhePheIle 252
    |||
    |||
    |||
Db 1411 -----AGTGTAGAGCTGAATCCGAAAGCTGCAGTACTATCT 1449
Qy 253 SerIyTySerAsnGlnLysAsnProHis-----Cys-----ValSer 265
    |||
    |||
    |||
Db 1450 GTGTGATTCAGTAAAGAGCGCAAGATATATACGTGAGATGTCCGCTCGTCTGCC 1509
Qy 266 LeuTyTyLysLeuSerSeryProGluAspAspProThrCysLysThrLysGluPheTrpAla 285
    |||
    |||
    |||
Db 1510 CTCTATACTCTACACAGCGCTGATATAAAGGCTGAGAGCTCTGAGAC---AAT 1566
Qy 286 ThrIleLeuAspSeryAlaGlyProLeuProAspTyThrProProGlu-----Ile 302
    |||
    |||
    |||
Db 1567 TCAGCTTGGAAATAATG-----CTGCAGAAATGCCAGATGCCCTCCAAAAAAGCTGAC 1620
Qy 303 PheSerPheGluSerThrThrGlyPheThrLeuTyGlyMetLeuTyLysProHisAsp 322
    |||
    |||
    |||
Db 1621 TTCAATTATTGGAATGAACAAATTT---TCGTATCAGATGATCTCTCCCTCAT---1674
Qy 323 LeuGlnProGlyLysLysTyProThrValLeuPheIleTyGlyGlyPro-----339
    |||
    |||
    |||
Db 1675 TTTGATTAATCCAGAAATATCTCTACTATTAGATGATGACAGGCCCATGATGCAA 1734
Qy 339 -----339
Db 1735 AAAGCAGACACTGTCTTGACAGCTGAAGTGGGCACCTTACCTTGCAAGCAGAAAAACATT 1794
Qy 339 -----339
Db 1795 ATAGTAGCTAGCTTGTATGCGAGAGAAAGTGTACCAAGAGATAGATCATGCATGCA 1854
Qy 339 -----339
Db 1855 ATCAACAGAGACTGGGAACTTTGAAGTTGAAGATCAAAATTGAAGCAGCCAGACAATTT 1914
Qy 339 -----339
Db 1915 TCMAAAATGGGATTTGTGGAACAACAAGAAATTGCAATTTGGGGCTGGTCATATGAGGG 1974
Qy 340 -----GlnValAlaIleAlaGly 345
    |||
    |||
    |||
Db 1975 TACGTAACTCAATGCTCTGGGATCGGGAAGTGGCGTGTCAAGTGTGGAATAGCCGTG 2034
Qy 346 AlaProValThrLeuTrpIlePheTyArgPheTrpGlyTyThrGluArgTyMetGly---364
    |||
    |||
    |||
Db 2035 GCGCTGTATCCCGGTGGAGTACTATGACTAGCTAGCTACAGAAAGCTTACATGGGTCTC 2094
Qy 365 ---HisProAspGlnAsnGluGlnGlyTyTyTyLysGlySerValAlaMetGlnAlaGlu 383
    |||
    |||
    |||
Db 2095 CCAACTCCAGAAAGACCTTGACCATTAACAAGAAATTCAAGTCATGAGCAGACTGAA 2154
Qy 384 LysPheProSeryGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 403
    |||
    |||
    |||
Db 2155 AATTTT-----AAACAAGTTGAGTACCTCTTATTTCATGAAACAGCAGATGATTAAGCT 2208
Qy 404 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyAsp 423

```

```

Db 2209 CACTTTCAGCAGTAGGCTCAGATCTCCAAAGCCTGGTTCAGTGTGAGATTTTCAG 2268
Qy 424 LeuGlnIleTyProGlnGluArgHisSeryIleArgValProGluSeryGluHisTy 443
    |||
    |||
    |||
Db 2269 GCATGTGTATCTAGTATGAAAGACCATGAAATACTGACGACAGCAGCAACATATTA 2328
Qy 444 GluLeuHisLeuLeuHisTyTyLeuGlnGlu 453
    |||
    |||
    |||
Db 2329 TATACCCATAGAGCCACTTCATTAACAA 2358
RESULT 5 0 4
PCT-US93-07923-1
; Sequence 1, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Tohichiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Frazer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR GEO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2924
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US93-07923-1
Alignment Scores:
Pred. No.: 4,05e-17 Length: 2924
Score: 234.00 Matches: 120
Percent Similarity: 33.10% Conservative: 70
Best Local Similarity: 20.91% Mismatches: 190
Query Match: 9.43% Indels: 194
DB: 5 Gaps: 25
US-10-070-464-5 (1-465) x PCT-US93-07923-1 (1-2924)
Qy 2 GlyThrAlaAsnProLysValThrPheLysMetSeryGluIleMetIleAspAlaGlu 21
    |||
    |||
    |||
Db 788 GAGAGCTGGAATCCAACTGTAAAGTTCTTTGTT-----GTAATATACAGACTCTTCAGC 841
    |||
    |||
    |||
Qy 22 ArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 41
    |||
    |||
    |||

```

```

Db      842 TCAGTCACCAATGCAATCTTCATACAAATCACTGCTCTCTATGTTGATA---GGG 898
Qy      42 ValGlyTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTrpSerIleLeu 61
Db      899 GATCACTACTGTGTGATGATGATGATGCAATGGCA----- 928
Qy      62 LeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIlePro 81
Db      929 -----ACACAAGAAAGAAATTTCTTTCAGGCGCTC----- 958
Qy      82 ValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThr 101
Db      959 -----AGGAGGATTCAGAACTAT-----TCGGTCAATG 985
Qy      102 ProeuIleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHis 121
Db      986 GATATTGTGACTATGATGAAATCCAGTGAAGATGG---AACTGC----- 1027
Qy      122 ValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLeuThr 141
Db      1028 TTAGTGGCAGCGCAACATGTAATGATGACTAGCTGGCTGGTTGGAAGATTTAGGCTT 1087
Qy      142 GlyPheArgHis-----LeuTyrLysIleThrSerIleLeuLys 154
Db      1088 TCAGAACCTCATTTTACCCCTTGATGTATAGCTTCTTACAAAGATCATCAGC-----AAT 1141
Qy      155 GluSerLysTyrLysArgSerSerGlyLysLeuProAlaProSerAspPheLysCysPro 174
Db      1142 GAAGAACGTTTACACACACATT-----TCTATTTTCCAAATATGAT 1180
Qy      175 IleLysGluGluIleAlaIleThrSerGlyLysTrpGluValLeuGlyArgHisGlySer 194
Db      1181 AAAAAAGACTGCACATTATTATTAACAAAAGGACCTGGGAAGTCATCGGG----- 1228
Qy      195 AsnIleGlnValAspGluValArgArgLeuValTyrPheGluGlyThr---LysAspSer 213
Db      1229 ---ATAAGAGCTTAACCAAGATTAATCTATATCACTTAACTAAATGAAATAAAGAAATG 1285
Qy      214 ProLeuGluHisIleLeuTyrValIleValSerTyrValAsnProGlyGluValThrArgLeu 233
Db      1286 CCAGAGAGAGAGAGATCTTTATAAATC-----CAACTT 1318
Qy      234 ThrAspArgGlyTyrSerHisSerCysCysIleSer-----GlnHisCys 248
Db      1319 AGTGAC-----TATACAAAAGTGACATGCTCAGTTGTGAGTGAATCCGGAAGAGTGT 1372
Qy      249 AspPheHeIleSerLysTyrSerAsnGlnLysAsnProHis-----Cys----- 263
Db      1373 CAGTACTATTCTGTGTCATTCAAGTAAAGGCGAAGTATATCAGCTGAGATGTTCCGCT 1432
Qy      264 -----ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLys 281
Db      1433 CCTGGTGTCCCTCTTACTTACTTACACAGCAGCTGAATGATAAAGGCTGAGAGTCTTG 1492
Qy      282 GluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGlu 301
Db      1493 GAAGAC---AATTCAGCTTGGATTAATG---CTGCAAGATGTCGAGATGCCCTCC 1543
Qy      302 -----IlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 318
Db      1544 AAAAACTGACCTCATTAATTTGAAAGAAACAAATTT---TGGTATCAGATGATCTTG 1600
Qy      319 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 338
Db      1601 CCTCCTCAT---TTGATTAATCCAAAGAAATATCCTTACTATTAAGTGTATGAGCG 1657
Qy      339 Pro----- 339
Db      1658 CCATGTAGTCAAAAAGACAGACTGTCTTCAAGTGAATGGGCACTTACCTTGCAAGC 1717
Qy      339 ----- 339
Db      1718 ACAGAAAACATTATAGTACTTGTGATGCGAGAGAAAGTGTTACCAAGAGATTAAG 1777

```

```

Qy      339 ----- 339
Db      1778 ATCATGCAATGCAATCACAAGACGTGGAAATTTGAAGTTGAAGATCAATTGAAGCA 1837
Qy      339 ----- 339
Db      1838 GCCAGACAATTTGAAAATGGGATTTGTGGACAACAAACGAATTCGAATTTGGGGCTGG 1897
Qy      340 -----GlnVal 341
Db      1898 TCATATGAGGGTACGTAACTCAATGCTCGGATGATCAGGAAGTGGCGTTCAAGTGT 1957
Qy      342 AlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspTrpGlyTyrThrGluArg 361
Db      1958 GGAATAGCGCTGGCGCGCTGATATCCCGGTGGAGTACTATGACTCAGTGTACACAAAGCT 2017
Qy      362 TyrMetGly-----HisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379
Db      2018 TACATGGGTCTCCCACTCCGAAAGACAACTTACCACTTACAGAAATTCAAACAGTCATG 2077
Qy      380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
Db      2078 AGCAGACTGAAAATTTT-----AAACAAGTTGAGTACCTCTTATTCATGGAACAGCA 2131
Qy      400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419
Db      2132 GATATATACGTTCACTTTCAGACAGCTCAGTCAATCTCCAAAGCCCTGGTGTGATGTTGA 2191
Qy      420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
Db      2192 GTGATTTCCAGGCAATGTGTATCTGATGAAACCAATGGAATTAAGCTAGACAGCAACGA 2251
Qy      440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlu 453
Db      2252 CACCAACATATATATACCCACATGAGCAGCATTCATAAAACAA 2293

RESULT 6
US-08-230-491A-1
Sequence 1, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELFE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```



LENGTH: 2815 Base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
-08-230-491A-1

Alignment Scores:  
Seq. No.: 8.88e-15 Length: 2815  
Score: 214.00 Matches: 87  
Percent Similarity: 32.48% Conservative: 53  
Percent Local Similarity: 20.19% Mismatches: 127  
Insert Local Similarity: 8.62% Indels: 164  
Insert Match: 1 Gaps: 13

S-10-070-464-5 (1-465) x US-08-230-491A-1 (1-2815)  
Y 140 LysThrGlyPheArgHisLeuTyrLysSerIleuLeuGluSerIleuLysGluSerLysLys 159  
b 1319 AAGGATGGCTACAAACATATCTACTATATC----- 1348  
y 160 ArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleuLysGluGlu 179  
b 1349 -----AAAGCACTGTGAAATGCTATT 1372  
y 180 AlaIleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199  
b 1373 CAATTTACAAATGCGCAAGTGGAGGCCATA-----AATATA----- 1408  
y 200 GluValArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleu 219  
b 1409 -----TTCAAGATTAACACAGGATTCACCTGTTTATTCTAGCAAT 1447  
y 220 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgLysTyrSer 239  
b 1448 GAATTTGAAGAATACCTCGGAGAGAGAAACATCTACAGATTAGCTTGAAGCTATCTCT 1507  
y 240 HisSer---CysCysIleSerGlnHis-----CysAspPhePheIleSer 253  
b 1508 CCAGGCAAGAGAGTGTCTTACTTGCCTTACGAAAGAAAGGTCGCAATTTACACAGCA 1567  
y 254 LysTyrSerGlnGlnLysAsnProHisCysValSerLeuTyr-----LysLeu 269  
b 1568 AGTTTCAGCGCTACGCCCACTATGCACTTGTCTGCGGCCGACGAGCATCCCATTT 1627  
y 270 SerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIleLeuAsp 289  
b 1628 TCACCCCTTCATGATGACGACATCATCAAGAA-----ATTAATCTCTGGA 1675  
y 290 SerAlaGlyProLeuProAsp-----TyrThrProGluIlePheSer 304  
b 1676 GAAACCAAGAAATTTGAAATATCTTTGAAAAATATCCAGCTGCTAAAGAGAAATTAG 1735  
y 305 PheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 324  
b 1736 AAACCTGAAGTAGAATAATTACTTTATGTAACAAGATGATTTCTTCTCTCAATTGAC 1795  
y 325 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro----- 339  
b 1796 AGATCAAGAGATATCCCTTGTATATCAAGTATGATGATGCTCTCCGACATGAGTGA 1855  
y 339 ----- 339  
b 1856 AGCTGTGATTGCTGTTAATGATATCTTATCTTCAAGTAGAAGAGAGTGCATT 1915  
y 339 ----- 339  
b 1916 GCCTGTGATGTCGAGAAACAGCTTCCAAAGTGACAAACCTCTTATGACGTAT 1975  
y 339 ----- 339  
b 1976 CGAAAGCTGGTGTATTATGAAAGTGAAGACAGATTACAGCTGTCAAAATTTACAGAA 2035  
y 339 ----- 339

Db 2036 ATGGCTTCATTGATGAGAAAAAGATAGCCATATGGCGCTGCTCTATGAGAGATACGTT 2095  
y 340 -----GlnValAlaIleAlaGlyAlaPro 347  
Db 2096 TCATCACTGGCCCTTGCAATCTGGAACCTGCTTTTCAAATGTGATATAGCAGTGGCTCA 2155  
y 348 ValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHisPro--- 366  
Db 2156 GTCTCCAGCTGCGAATATATACGCTCTCTCTACACAGAGATATCATGTCGCTCCACAA 2215  
y 367 ---AspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhe 385  
Db 2216 AAGATGATATATCTTGACACACATTAAGATTCATCTGATGATGCGCAAGACCAATATTTC 2275  
y 386 ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPhe 405  
Db 2276 AAGAAATGTAGAC-----TATCTTTCATCCACGGAACGACAGATATATATGACACTT 2329  
y 406 AlaIleThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425  
Db 2330 CAAACTCAGCACAGATGTGCTAAAGCTCTGTTAATGACACAGTGAATTTCCAGGCAATG 2389  
y 426 IleTyrProGlnGluArgHisSerIleArgValProGluSerGly-----GluHis 442  
Db 2390 TGGTACTCTGACCAAGAACCAAGCGCTTA-----TCCGCTCTGACAGAACCCAC 2437  
y 443 TyrGluLeuHisLeuLeuHisTyrLeuGlnGlu 453  
Db 2438 TTATACACCCACATGACCACTCTCTAAAGCAG 2470

RESULT 7  
US-08-619-280A-1  
Sequence 1, Application US/08619280A  
Patent No. 5767242  
GENERAL INFORMATION:  
APPLICANT: Zimmermann, Rainer; Park, John E.;  
TITLE OF INVENTION: RETIRED, Wolfgang; Old, Lloyd J.  
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN  
TITLE OF SEQUENCES: ALPHA, AND USUS THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Pelfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage  
COMPUTER: IBM PS/2 PC-DOS  
OPERATING SYSTEM: wordperfect  
SOFTWARE: wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/619,280A  
FILING DATE: 18-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/230,491  
FILING DATE: 20-APRIL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5767242man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5330.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 638-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2815 Base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-619-280A-1

**Alignment Scores:**

Pred. No.:	8.88e-15	length:	2815
Score:	214.00	Matches:	87
Percent Similarity:	33.48%	Conservative:	53
Best Local Similarity:	20.19%	Mismatches:	127
Query Match:	8.62%	Indels:	164
DB:	1	Gaps:	13

US-10-070-464-5 (1-465) X US-08-619-280A-1 (1-2815)

QY	140	LYSTHNGIYPHARGHISLEUYRVLVSIETHTSERILEULEUYGLUSERLYSTRYS	159
Db	1319	AAGATGCGCTACAAACATATTCCATATAC	1346
QY	160	ARGSERSEGLYLEUPROALAPROSERASPHELYSCYSPROILEYEGLUJULE	179
Db	1349	-----AAAGACACGTGGAAAAATGCTATT	1372
QY	180	ALAIETHSERGLYGLUTROGLUVALLEUGLYARGHISGLYSETHNILEGVALASP	199
Db	1373	CAAAATTCAAGTGGCAAGTGAGGCCATA-----AAATA-----	1408
QY	200	GIUVALARGARGLUVALTYRPHHEGLUJULYTHLYASPSPROLEUGLHSHISEU	219
Db	1409	-----TTCCAGATGTAACACAGATTCCCTGTTTAATTACCAAT	1447
QY	220	TYRVALVLSERTYRVALASPPROGLYGLUVALTHARGLEUTHASPARGLYTRYSER	239
Db	1448	GAATTGGAAGAAATCCCTGGAAAGAAAGAAACATCTACAGATTAGCATTTGAAGCTATCT	1507
QY	240	HISER---CYSQYAILSESGLNHS-----CYASPHEPHEHLESER	253
Db	1508	CCAAAGCAAGAGTGTGTACTTGCCATCTAAAGAAAGAAAGTGCCATATTACACAGCA	1567
QY	254	LYSTYRSARENGILYASNPROIHISCYVALSERLEUYR-----LYLEU	269
Db	1568	AGTTTCACGACCTACGCCAAGTACTACGACTTGCTGTCTACGCCGCCAGCATCCCATTT	1627
QY	270	SESERPROGLUASPASPPROTHRCYSLYSTHRLYSGIUPHETPLATHRIELEUASP	289
Db	1628	TCCACCCCTTCATGATGAGCGACGTATCAAGA-----ATTAAATCCTGAA	1675
QY	290	SERLAGLYPROLEUPROASP-----TYRTHPROPROGLUILEPHESER	304
Db	1676	GAAGAACAGAAATGGAAAAATGCTTTGAAAAATATCCAGCTGCCTAAAGAGAAATTAAG	1735
QY	305	PHEGIUSERTHRHNGIYPHETHLEUYRGLYELUETHEUYRVLSPROIHISAPLEUGLN	324
Db	1736	AAACTTGAGTAGATGAAAAATTACTTTAAGGTACAAAGATTCCTTCCTCAATTGAC	1795
QY	325	PROGLYLSYSTRYPROTHRVALLUPHEHLEUYRGLYGLYPRO-----	339
Db	1796	AGATCAAGAAAGATACCTCTGCTAATTCAAGTGHATGSGTCCCTGCAGTCAGAGTGTA	1855
QY	339	-----	339
Db	1856	AGTCTGTATTGCTGTTAATTGATATCTATCTTGCAAGTAAGAAAGGATGTCATT	1915
QY	339	-----	339
Db	1916	GCCTTGTGATGATGTCGAGAACAGCTTCCAAAGTGCANAACTCTTATGACGTAT	1975
QY	339	-----	339
Db	1976	CGAAAGCTGGGTGTTTATGAAGTTGAAGACCAAGATTACAGCTGCAGAAAAATTCATGAA	2035
QY	339	-----	339
Db	2036	ATGGGTTTCATTGATGAAAAAAGATAGCATATGGGGCTGGTCCCTATGAGAGTACGTT	2095
QY	340	-----GLNVALAIAIEALGYALAPRO	347

```

Db      2096  TCATCACTGAGCCCTTGATCGATCGAACTGGTCTTTTCAATAATGTGATAGCAGTGGCTCCA      2155
QY      348  ValThLeuTrPlllePheTyrAspThrGlyTyrThrGluArgTyrMeGlyHisPro--- 366
Db      2156  GTCTCCACTGGAGATATATTACGCCGTCTGTCTAACACAGAGAAATTCATGGGCTCCCAACA      2215
QY      367  ---AapGlnAaNGluGlnGlyTyrTyrLeuGlySerValAlaMeGlnAlaGluLysPhe      385
Db      2216  AAGAGTAGTAAATCTTGAGACCACTATAGAATTCAACTGTGATGCGAAGACGCAAAATATTTC      2275
QY      386  ProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAapGluAsnValHisPhe      405
Db      2276  AGAAATAGAGAC-----TATCTTCTCATCCAGGAAACAGACAGATGATATGTGCACCTT      2329
QY      406  AlaHisThrSerIleLeuLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln      425
Db      2330  CAAACTGACGACAGATTGCTTAAAGCTGTGCTTAATGACAAAGTGAATTTCACAGCAATG      2389
QY      426  IleTyrProGlnGluArgHisSerIleArgValProGluSerGly-----GluHis      442
Db      2390  TGTGATCTGTGACCAAGAACCGGCTTA-----TCCGCGTGTCCACGAACAC      2437
QY      443  TyrGluLeuHisLeuLeuHisTyrLeuGlnGlu 453
Db      2438  TTATACACCCCATGACCATGCTCTTAAGACAG 2470

RESULT 8
US-08-940-391-1
; Sequence 1, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rüttig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-940-391-1

```

**Alignment Scores:**

Pred. No.:	8.88e-15	2815
Score:	214.00	87
Percent Similarity:	32.488	Conservative: 53
Best local Similarity:	20.194	Mismatches: 127
Query Match:	8.62%	Indels: 164
DB:	2	Gaps: 13

US-10-070-464-5 (1-465) x US-08-940-391-1 (1-2815)

Oy	140	YysThrGlyPheAsgHisLeuTyrTyrLeuTyrSerIleLeuLysGluSerTyrTyrLys	159
Db	1319	AAGGATGGCTCAAAATCAATTCACATATTC	1348
Oy	160	ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIle	179
Db	1349	-----AAAGACATCGTGAAGAAAGCTATT	1372
Oy	180	AlaIleThrSerGlyGluTyrGluValLeuGlyAsgLysSerAsnIleGluValAsp	199
Db	1373	CAAAATTCAAGTGGCAAGTGGGAGGCCATA-----AAATATA-----	1408
Oy	200	GluValAsgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleu	219
Db	1409	-----TTCAAGTACACAGGATTCACGTTTATTATTCACAAAT	1447
Oy	220	TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer	239
Db	1448	GAATTGGAAGATACCCCTCGAAGAAACAATCTACAAATTAAGATTGGAAGCATCTCT	1507
Oy	240	HisSer---CysCysIleSerGlnHis-----CysAspPhePheIleSer	253
Db	1508	CCAAGCAAGAAGTGGTACTTGCCATCTTAAGAAAGAAAGAGTGCATATTACACAGCA	1567
Oy	254	LysTyrSerAsgGlnLysAsnProHisCysValSerLeuTyr-----LysIleu	269
Db	1568	AGTTTCACGCACTACGCCCAAGTACTATGACATTGTCGTACGGCCACGCAATCCCATTT	1627
Oy	270	SerSerProGluAspAspProThrCysLysThrLysGluPheTyrAlaThrIleLeuAsp	289
Db	1628	TTCACCCCTTCAATGATGAGACGCACTGATCAAGA-----ATTAAATCTCGAA	1675
Oy	290	SerAlaGlyProLeuProAsp-----TyrThrProProGluIlePheSer	304
Db	1676	GAAGAACAGAAATGGAAAAATGCTTTGAAAAATATCCAGCTGCGCTTAAGAGAAATTAAG	1733
Oy	305	PheGluSerThrThrGlyPheThrIleuTyrGlyMetLeuTyrLysProHisAspLeuGln	324
Db	1736	AAACTTGAAGTAGATGAATTAATCTTAAAGTAGACAAAGATCTTCCCTCAATTGGAC	1795
Oy	325	ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro-----	339
Db	1796	AGATCAAGAAAGATATCCCTTGCTAATTCAAGTATATGGTGGTCCCTGACATCAGAGTGA	1855
Oy	339	-----	339
Db	1856	AGGTCTGATTGCTGTAAATTGATATCTTATCTTGAAGTAAGAAAGAGTGGTCATT	1915
Oy	339	-----	339
Db	1916	GCTTGTGTGATGTGCGAGAACAGCTTCCAAAGGTGCAAACTCTCTATGACAGTAT	1979
Oy	339	-----	339
Db	1976	CGAAAGCTGGGTGTTATGAAGTTGAACACAGATTACAGCTGTCAGAAAAATTCATAGAA	2035
Oy	339	-----	339
Db	2036	ATGGGTTTCATTGATGAAGAAAAAAGATAGCCATATGCGGCTGATCTTATGAGAGATAGCTT	2095
Oy	340	-----GlnValAlaIleAlaGlyValPro	347
Db	2096	TCATCACTGGCCCTTGATCTTGAACTGAGTCTTTTTCAAATATGTGATTATGACGTGGCTCA	2155

Oy	348	ValthrlleuTPrIlePheTyAspThglYlyrThrGluAGlyrMeGlYHisPro---	366
Db	2156	GTCCTCAGCTGGGAATATTAGCGCTGTCTTACACAGAGAAATTATCGGCTTCCCAACA	2215
Oy	367	---AepGlnaengIugInglYTyTyrcIeugIySerValAlametGlnalaglulysPhe	385
Db	2216	AAGATGATTAATCTTGAGCACTATAAAGATTCAACTGGATGGCAGAAGCGAATATTTC	2275
Oy	386	ProSerGluProAnaArgLeuLeuLeuHisglYPheLeuAepGluAsnValHisPhe	405
Db	2276	AGAAATGTAAGC-----TATCTTCTCATCCACGGAACAGCATGATGATATGTGCATTT	2329
Oy	406	AlaHisThrSerIleLeuLeuSerPheLeuValAlarglaglYlyProTyAspLeuGln	425
Db	2330	CAAACTCAGCACGATGCTTAAAGCTCTCGGTTATGACACAAAGTGGAATTCACGCAATG	2389
Oy	426	IleTyTrProGlnGluArgHisSerIleArgValProGluSerGly-----GluHis	442
Db	2390	TGGTACTCTGACCGAACCCACGAGGCTTA-----TCGGCGCTGTCCACGAACCAAC	2437
Oy	443	TyrGluLeuHisIleLeuLeuHisTyTyrcIeugInlu	453
Db	2438	TTATACACCCACATGATCCCACTTCTCTTAAGCAG	2470

## RESULT 9

US-08-699-103B-1/c  
; Sequence 1, Application US/08699103B

; GENERAL INFORMATION:

APPLICANT: Hampton, Randol

**TITLE OF INVENTION: CHOLESTEROL SYNTHESIS**

CORRESPONDENCE ADDRESS:

**STREET: 2200 Sand Hi**

STATE: CA

COUNTRY: USA  
ZIP: 94025

MEDIUM TYPE:

OPERATING SYSTEMS  
COMPUTER: IBM  
; ;

SOFTWARE: CITIBENT ADDIT

APPLICATION NUMBER: US/08/699,103B  
FILING DATE: 15 AUG 2006

PRIOR APPLICATION DATA:

FILING DATE: 17-AUG-1995

NAME: Green, Grant D.

REFERENCE/DOCKET NUMBER: 09272/005001

TELEPHONE: 650/322-5070

; INFORMATION FOR SEQ ID NO: 1

LENGTH: 4982 base pairs

STRANDEDNESS: single

MOLECULE TYPE: Genomic DNA

8  
 8  
 5  
 5  
 4  
 4  
 4  
 4  
 4

6.98e-12	Pred. No.:
6.98e-12	6.98e-12

Percent Similarity: 40.97%

Alignment Scores:	
Pred. No.:	6.98e-12
Score:	193.00
Percent Similarity:	40.97%
	Length: 4962
	Matches: 65
	Conservative: 28



Oy 390 AsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAenValHisPheAlaHisThrSer 409  
Db 4470 AATGATTTTGTGATCGACGGAACAGAGATGATTAAGTTCACTTCAAAATTCCTTA 4411  
Oy 410 IleLeuLeuSerPheLeuValArgAlaGly--LysProTyrAspLeuGlnIleTyrPro 428  
Db 4410 AAGTTTCTGGACCTTTTGATCTTAATGTTGTGGAATAATTAAGACGTCCAGCTTTCTCT 4351  
Oy 429 GlnGluArgHisSerIleArg 435  
Db 4350 GACTCAGATCATAGTATTAAGA 4330

RESULT 11  
US-09-628-133-1/c  
Sequence 1, Application US/09628133  
Patent No. 6531292

GENERAL INFORMATION:  
APPLICANT: Rine, Jasper D.  
APPLICANT: Hampton, Randolph  
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
NUMBER OF INVENTIONS: CHOLESTEROL SYNTHESIS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 2200 Sand Hill Road, Suite 100  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/628.133  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/699,103  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D.  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 09272/005001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/332-5070  
TELEFAX: 650/854-0875  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4982 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-09-628-133-1

Alignment Scores:  
Pred. No.: 6.98e-12 Length: 4982  
Score: 193.00 Matches: 65  
Percent Similarity: 40.97% Conservative: 28  
Best Local Similarity: 28.63% Mismatches: 80  
Query Match: 7.78% Indels: 54  
Gaps: 9

US-10-070-464-5 (1-465) x US-09-628-133-1 (1-4982)

Oy 244 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAenGlnIlyAsnProHisCys 263  
Db 4953 TTAAGTGCACCACTATCCCTGTTATTTCTTTCGATATGGGGACCGAATTTCAACAAAGTT 4894  
Oy 264 ValSerLeuTyrLysLeuSerSerProGlu----- 273  
Db 4893 GTCAAAAGCTTTCCGTAGATTAATGATGAGTGTAGCTTCACAAATTAAACGCAATTGTA 4834

Oy 274 -----AspAspProThrCysLysThrLysGluPheTyrAlaThrIleLeu 288  
Db 4833 GTTGTGTTGACCGGTGCTGCTGACTGCGCTCAAAAGTCAAGACTTTAGATCCCTGTTGCGC 4774  
Oy 289 AspSerAlaGlyProLeuProAspTyrThrProGluIlePheSerPheGluSerThr 308  
Db 4773 GATAGGCTCGGT-----GATTACGAGGCCCGCCCAAAATATGCGCGCTTCCTTA 4723  
Oy 309 ThrGly-----PheThrLeuTyrGlyMetLeuTyr--- 318  
Db 4722 TATGTTCTTAACTTTGTTGATCCGAAAAGATTTCTTATTTGTTGTTGTCATACGGG 4663  
Oy 319 -----LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeu 333  
Db 4662 CGGTACCTGCACACATAAACTTTGAGAGAAAGATGCGGAAGACAT----- 4618  
Oy 334 PheIleTyrGlyGlyProGlnValAlaIleAlaGlyAlaProValThrLeuTyrIlePhe 353  
Db 4617 TTCAAAATACGGG-----ATGTCAGTTGCGCCAGTAACGAGCTGAGATTT 4573  
Oy 354 TyrAspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAenGluGlnGlyTyr 373  
Db 4572 TACGATTCCTGTTTAACTGAGAGGTACATGATCTCTCAAGAAAACTTTGATGATGATAC 4513  
Oy 374 TyrLeuGlySerVal-----AlaMetGlnAlaGluLysPheProSerGluPro 389  
Db 4512 GTAGATCAACAGCTTCAATATGTCATGCTTTG-----GCACAGCA 4471  
Oy 390 AsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAenValHisPheAlaHisThrSer 409  
Db 4470 AATGATTTTGTGATCGACGGAACAGAGATGATTAAGTTCACTTCAAAATTCCTTA 4411  
Oy 410 IleLeuLeuSerPheLeuValArgAlaGly--LysProTyrAspLeuGlnIleTyrPro 428  
Db 4410 AAGTTTCTGGACCTTTTGATCTTAATGTTGTGGAATAATTAAGACGTCCAGCTTTCTCT 4351  
Oy 429 GlnGluArgHisSerIleArg 435  
Db 4350 GACTCAGATCATAGTATTAAGA 4330

RESULT 12  
US-09-280-116-172  
Sequence 172, Application US/09280116A  
Patent No. 6331427

GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Pro tease Homologs  
FILE REFERENCE: 5800-24, 035800/176965  
CURRENT APPLICATION NUMBER: US/09/280,116A  
CURRENT FILING DATE: 1999-03-26  
NUMBER OF SEQ ID NOS: 268  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 172  
LENGTH: 502  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: prolyl oligopeptidases

US-09-280-116-172

Alignment Scores:  
Pred. No.: 1.89e-10 Length: 502  
Score: 167.00 Matches: 32  
Percent Similarity: 55.68% Conservative: 17  
Best Local Similarity: 36.36% Mismatches: 37  
Query Match: 6.73% Indels: 2  
Gaps: 1

US-10-070-464-5 (1-465) x US-09-280-116-172 (1-502)

Oy 346 AlaProAlaThrLeuTyrPheIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHis 365  
Db 346 AlaProAlaThrLeuTyrPheIlePheTyrAspThrGlyTyrThrGluArgTyrMetGlyHis 365

```

Db      67 GCACCTATCAGACCTGAATTGTATGCTTCAGCTTTCTGAAAGATACCTTGGGATG 126
QY      366 ProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlnLysPhe 365
Db      127 CCATCTTAAGAGAAAGACCTTACGAGCCAGGTGTGCTACATAATGTTCAATGCGCTTG 186
QY      386 ProSerGlnProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPhe 405
Db      187 AAGAAAGAA-----AATATATTAAATATCATGAGACCTGACACAAAGTTTCAATTC 240
QY      406 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425
Db      241 CAACACTGACGACAAATTAATCAACGACCTTAATAAGCTGAGTAATATATCTATGTCAG 300
QY      426 IleTyrProGlnGlnArgHisSer 433
Db      301 GTCTACCCAGATGAAGCTCATTAAC 324

RESULT 13
US-09-221-017B-962
; Sequence 962, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/0221, 017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 962:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 815 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS

```

```

; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...815
; US-09-221-017B-962

Alignment Scores:
Pred. No.: 1,12e-07 Length: 815
Score: 146.50 Matches: 67
Percent Similarity: 37.33% Conservative: 42
Best Local Similarity: 22.95% Mismatches: 114
Query Match: 5.90% Indels: 69
DB: 4 Gaps: 14

US-10-070-464-5 (1-465) x US-09-221-017B-962 (1-815)

QY      43 GlnTyrIleAlaArgAlaGlyTyrPheProGlnGlyLysTyrAlaTrpSerIleLeuLeu 62
Db      123 AAATTTCTGACGAATTTGAGTTGAGTCCGACGAAATAATCTTGATGTAGTGAAGTGTG 182
QY      63 AspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGlnLeu-----PheIle 80
Db      183 AATCGTGTCTCAAAACGAATGTAAAGTAAATGCTTATGACGCTGACACCGGTAGATTCGTC 242
QY      81 ProValGlnAspAspValMetGlnArgGlnArgLeuIleGlnSerValProAspSerVal 100
Db      243 -----CGTACGCTTTTGTGTAACCGATAACATTATGTA 278
QY      101 ThrProLeuIleIleTyrGlnGlnThrAspIleTrpIleAsnIleHisAspIlePhe 120
Db      279 GACCCGCTTA-----CATCCCTGACA 299
QY      121 HisValPheProGlnSerHisGlnGlnIleGlnPheIlePheAlaSerGlnCysLys 140
Db      300 TTCCTT---CCGGAAAGTACAAAT-----CAGTTCAATTGGCAGACCGCT---CGC 344
QY      141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGlnSerLysTyrLysArg 160
Db      345 GACGAGTGAACCATCTCTTCTGTATGAT----- 374
QY      161 SerSerGlyLeuProAlaProSerAspPheLysCysBroIleLysGlnIleAla 180
Db      375 ACTACAGTGTCTGTG-----ATCCGTGAC----- 398
QY      181 IleThrSerGlyGlnTyrProGlnValLeuGlyArgHisGlySerAsnIleGlnValAspGln 200
Db      399 GTGACAAAGAGGAGTGGAGGTTACAAACTTTGACGCG-----TTGCATCCC 446
QY      201 ValArgArgLeuValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleLeuTyr 220
Db      447 AAGGAAACACGGCTCTATTTCGAAAGTACCGAAGCCGCTCTCGAAGCCCATTTTTC 506
QY      221 ValValSerTyrValAsnProGlnGlnValThrArgLeuThrAspArgGlyTyrSerHis 240
Db      507 TGTATTGAT---ATCAAGAGAGAAAGACAAAGATCTGACTCCGGAGTGGGAATGCAC 563
QY      241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260
Db      564 CGCACTCACTATCTCTGATGAGTTCGCCATATAGATATTTTACAGTCACTACTGTCTC 623
QY      261 ProHisCysValSerLeuTyrLysLeuSerSerProGlnAspArgProThrCysLysThr 280
Db      624 CCGCGTAAGTTACAGTGAACAAATATCGGCAAGGCTTCAC----- 665
QY      281 LysGlnPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300
Db      666 -----ACACTCTTGAGAGGCTAAGAACCCCGATACGGGCTATGCGACGCG 710
QY      301 GlnIle-----PheSerPheGlnSerThrThrGlyPheThr---LeuTyrGlyMetLeu 317
Db      711 GAGATCAGAACGGGTACATCATCATGCGGCGCATGGGCGACACCTTTATTATCAAGCTC 770
QY      318 TyrLysProHisAspLeuGlnProGlnLysLysTyr 329

```

Db 771 ACGATGCCGCTTCATTTGATCCGGCAAGAAATAT 806

RESULT 14  
US-09-016-080-2  
Sequence 2, Application US/09016080  
Patent No. 6133012  
GENERAL INFORMATION:  
APPLICANT: Ishikawa, Kazuhiko  
APPLICANT: Matsui, Ikuo  
APPLICANT: Ishida, Hiroyasu  
APPLICANT: Kosugi, Yoshiyugu  
APPLICANT: Higuchi, Katsuhiko  
TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING  
FILE REFERENCE: 07898/022001  
CURRENT APPLICATION NUMBER: US/09/016,080  
EARLIER APPLICATION NUMBER: JAPAN 18381/1997  
EARLIER FILING DATE: 1997-01-31  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1896  
TYPE: DNA  
ORGANISM: Pyrococcus horikoshii  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1896)  
US-09-016-080-2

Alignment Scores:  
Pred. No.: 3,5e-06 Length: 1896  
Score: 139.00 Matches: 113  
Percent Similarity: 30.58% Conservative: 65  
Best Local Similarity: 19.42% Mismatches: 172  
Query Match: 5.60% Indels: 232  
DB: Gaps: 28

US-10-070-464-5 (1-465) x US-09-016-080-2 (1-1896)

QY 7 LysValThrPheIysMetSerGluIleMetIleAspAlaGluGlyArgIleIleAspVal 26  
Db 307 AAGTTCAATAACGGGATTAAGAAACCTCGCTTACCGAGATGGGAAAGTATAGCCGTG 366  
QY 27 ILe-----AspLysGluLeuIleGlnProPhe 35  
Db 367 GTTACCCCTATAGACGTTGAGAAAAGGAGATGATACGTTCACTATATAGGGAATA 426  
QY 36 GluIleLeuPheGluGlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLys 55  
Db 427 CCATTCTGTGTTTATAGAGTT-----GGCTGGATC--TAGGGAATA 465  
QY 56 TyrAlaTyrSerIleLeuLeuAsp-----ArgSerGlnThrArgLeu----- 69  
Db 466 AGAAACCTGTCTACCTGTGAGCTTGAGAGCGGGAAGAAAAGACATTAATCTCCAAAG 525  
QY 70 -----GlnIleValIleSerProGluLeuPheIleProValGluAsp 84  
Db 526 AACCTAAATGTGATCGATAGAGTTCCACACGCGTACATATACCTCAGGCCCAAGAG 565  
QY 85 Asp-----ValMetGluArgGln----- 90  
Db 586 GATAGGGAAGAAACCTCTGATATCCGATCTTACCTCTCGACGATATAGAAAGTTAGG 645  
QY 91 -----ArgLeuIleGluSerValProAspSerValThrProLeu 103  
Db 646 AAGCTGACCCCAAGGAAGTGAAGGATACCTGACCTCTCCCTTATATGACGGAAGCTTC 705  
QY 104 IleIleTyrGluGluThrThrAsp-----IleTyrIleAsnIleHisAspIlePheHis 121  
Db 706 GTACTTAAGGCTAAACCTTATAGAAAGGGAATCCCAACCAAGCCGCAC--ATCTACAC 762  
QY 122 ValPheProIleSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThr 141

Db 763 TAGCATCCC----- 777  
QY 142 GlyPheArgHisLeuTyrIlySerIleThrSerIleLeuLysGluSerLysTyrLysArgSer 161  
Db 778 GGA-----GAACTTAAGAGCTCACAAAGATTTAGACAGAAAGCTTACAACTCTTA 831  
QY 162 SerGlyGlyLeuPheProAlaProSerAspPheLysCysAspProIleLysGluGluIleAlaIle 181  
Db 832 AAC-----TCCGATGTTCCAGGAAGTACAGAGCGCCGAGCTGTGTAC 873  
QY 182 ThrSerGlyGluTyrPheValLeuGlyArgHisGlySerAsnIleGlnValAspGluVal 201  
Db 874 AAGGAGGGG--TGG----- 885  
QY 202 ArgArgLeuValTyrPheGluGluGlyThrLysAspSerProLeuGluHisLeuTyrVal 221  
Db 886 -----ATCTACTATGTCCGACG--GATGGCCCT--AGGCAAACTCTTTAGG 930  
QY 222 ValSerTyrValAsnProGlyGluValThrArgLeuThr-----AspArgIlyTyrSer 239  
Db 931 GTCACTTA-----GATGAAAGATTTGAAGGGTATAGTGGAAGTATAGAGCTTGAA 984  
QY 240 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259  
Db 985 -----ACCTCGATATAGGGGATTAATACATAGCTTTCACG 1017  
QY 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLys 279  
Db 1018 GCTCAAGATGCTGTAAACCCCACTGATATACATATACAGGATGGAAGAAAGAAAG 1077  
QY 280 ThrLysGluPhe-----TyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyr 297  
Db 1078 GTTACGACGCTTAAACAATGG-----ATTAAGCGTTAC 1110  
QY 298 Thr-----ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 315  
Db 1111 ACCCTTCAAAACCTGAACACTTAAAGTTAAAGCAAGTACGCGGGTTGAATATAGATGCC 1170  
QY 316 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 335  
Db 1171 TGGGTATATGAACCGGTGAACCTTCAGAGAAAGAAAGATATCCAGCTATTTAGACATC 1230  
QY 336 TyrGlyGlyProGluValAlaIleAla----- 344  
Db 1231 CACGGTGGTCTTAACCGCTTACGGTTACGCTTTATATGACAGAGTTCCACGTTTAAAC 1290  
QY 344 ----- 344  
Db 1291 TCTAAAGGCTTCGTCGTGATATTCTCAAACTTAGAGAGAGCGATGGCTAGGAGAGAG 1350  
QY 344 ----- 344  
Db 1351 TTCGGGATATTAAGGGAGACCTATGGGAGAGGAGATTTACGAGATTTATAGAGTATGC 1410  
QY 344 ----- 344  
Db 1411 GATGAAGCATTAAGAGATTTGACTTCATATAGTGGGAAAAGCTAGAGATCCGGCGGT 1470  
QY 345 -----GlyAlaProValThrLeuTyrPheThrLysAspThrGlyTyrThrGluArgTyr 362  
Db 1471 TCCATATGGTGGCTTCATATGACGAACTGATA-----GTCCGACATATCAACAGGTTCC 1521  
QY 362 ----- 362  
Db 1522 AAAGCGCGCTTAACCCAGAGATCAATTTCAATTGCAATAGCTTTTCGGGACAACGAGAT 1581  
QY 363 MetGlyHis-----ProAspGln-----AsnGluGlnGly 372  
Db 1582 ATAGTTATATCTTCTGTCAGATCAATATAGAAAGAGATCCCTGAGCAACTTGAAAGGT 1641  
QY 373 TyrTyrLeuGlySerValAlaMetGlnAlaGluLysPheProSerGluProAsnArgLeu 392



Db 1642 TATTGGGAAAAAGCCATTAAAGTAGCT-----CCCAACGTTGAAACTCCCTG 1692  
Qy LeuLeuLeuHISGLYPhLeuLeuSpGLuSnValHISpHeAlaHISThrSerIleLeuLeu 412  
|||:|||||  
Db 1693 CTTATATCCACTCTACCGAAGACTACAGGCTTGGCTCCCGAGGCACTTGCACTCTTC 1752  
Qy 413 SerPheLeuValArgAlaGLYLYSPProTYrAspLeuGlnIleTYrProGlnGluArgHIS 432  
|||:|||||  
Db 1753 ATATCCCTTAAATACCTGGGAGAGAGATTGATTCGCAATATTCCTCCAGAGAAAAATCAT 1812  
Qy 433 SerIle 434  
1813 GACCTA 1818  
RESULT 15  
US-09-221-017B-253  
Sequence 253, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEO ID NO: 253:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 543 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYBRIDICAL: NO  
ANTI-SENSE: UNKNOWN  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1...543  
US-09-221-017B-253

Alignment Scores:  
Pred. No.: 2,616-06 Length: 543  
Score: 132.50 Matches: 30  
Percent Similarity: 51.69% Conservative: 16  
Best Local Similarity: 33.71% Mismatches: 40  
Query Match: 5,34% Indels: 3  
DB: 4 Gaps: 1  
US-10-070-464-5 (1-465) x US-09-221-017B-253 (1-543)  
Qy 340 GlnValAlaIleAlaGLYAlaProValThrLeuTrpIlePheTYrAspThrGLYTYrThr 359  
|||:|||||  
Db 283 AAAGTCGAGTAGCGCGGCGCTGTCTATAGACTGGAATGCAATGAGATTATGACGCT 342  
Qy 360 GlnArgTYrMetGlnHISProAspGLuSnGlnGlnGLYTYrTYrLeuGLYSerValAla 379  
|||:|||||  
Db 343 GAGCGTTATTTTCGATGCCGCCACAGGAAAAATCCCGAAGATACGATGCTGCCAACCTGCTC 402  
Qy 380 MetGlnAlaGLYLYSPProSerGLuProAsnArgLeuLeuLeuHISGLYPhLeu 399  
|||:|||||  
Db 403 AAAGCGCGCGGTATCTG-----AAAGCAGCACTTATGCTATTCATGAGCGATC 453  
Qy 400 AspGlnSnValHISpHeAlaHISThrSerIleLeuLeuSerPheLeuValArgAlaGLY 419  
|||:|||||  
Db 454 GATCCGCGTCGTGTATGCGACATTCATCTTTCTTGTATGCTTGCGTGAAGGACGCG 513  
Qy 420 LysProTYrAspLeuGlnIleTYrPro 428  
|||:|||||  
Db 514 ACCTATCTGACTCTTACGTCTATCCG 540

Search completed: October 16, 2003, 03:31:00  
Job time : 100.851 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 16, 2003, 00:05:32 (without time 347.655 seconds  
(without alignments)  
3512.533 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTATPKVTFKMSIMDAE.....HLHYQENLGSRIALXVI 465

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1333063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US10070464/runat\_15102003\_113555\_24924/app\_query.fasta\_1.2652  
-DB=Published Applications NA -QMT=fastcap -SUFFIX=trpb -MIMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USRR=US10070464 @CGN\_1.1 845 @runat\_15102003\_113555\_24924  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*

1:	/cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubna/PT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubna/PTCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubna/US09_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/2/pubna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/2/pubna/US09C_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq:*
16:	/cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:*
17:	/cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2446.5	98.6	4523	10	US-09-976-674-8 Sequence 8, Appli

2	2422	97.6	2649	12	US-10-054-776-1	Sequence 1, Appli
3	2422	97.6	2649	12	US-10-170-789-39	Sequence 39, Appli
4	2422	97.6	2671	10	US-09-976-674-2	Sequence 2, Appli
5	2422	97.6	3143	12	US-10-170-789-37	Sequence 37, Appli
6	2418.5	97.4	4676	10	US-09-976-674-20	Sequence 20, Appli
7	2402	96.8	4829	10	US-09-976-674-12	Sequence 12, Appli
8	2149	86.6	4685	10	US-09-976-674-22	Sequence 22, Appli
9	1494.5	60.2	2617	10	US-09-976-674-28	Sequence 4, Appli
10	1494.5	60.2	4219	10	US-09-976-674-28	Sequence 28, Appli
11	1494.5	60.2	4302	10	US-09-976-674-24	Sequence 24, Appli
12	1445	58.2	4180	10	US-09-976-674-36	Sequence 36, Appli
13	1445	58.2	4263	10	US-09-976-674-34	Sequence 34, Appli
14	1386	55.8	4309	10	US-09-976-674-14	Sequence 14, Appli
15	1342.5	54.1	4076	10	US-09-976-674-32	Sequence 32, Appli
16	1340.5	54.1	4159	10	US-09-976-674-30	Sequence 30, Appli
17	1296	52.2	3262	13	US-10-098-841-83	Sequence 83, Appli
18	1293	52.1	4037	10	US-09-976-674-40	Sequence 40, Appli
19	1293	52.1	4120	10	US-09-976-674-38	Sequence 38, Appli
20	1276	51.4	2801	13	US-10-098-841-100	Sequence 100, App
21	471.5	19.0	2079	13	US-10-044-090-843	Sequence 843, App
22	471.5	19.0	2411	10	US-09-976-674-26	Sequence 26, Appli
23	401.5	16.2	561	11	US-09-764-891-877	Sequence 877, App
24	291	11.7	4797	11	US-09-764-891-7074	Sequence 7074, App
25	279	11.2	502	11	US-09-918-995-19585	Sequence 19585, A
26	267	10.8	281	9	US-09-867-550-987	Sequence 987, App
27	249.5	10.1	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
28	249.5	10.1	4835	14	US-10-165-603-5	Sequence 5, Appli
29	247	10.0	2388	12	US-09-870-133-3	Sequence 3, Appli
30	247	10.0	2388	14	US-10-160-501-6	Sequence 6, Appli
31	247	10.0	2583	10	US-09-976-674-6	Sequence 1, Appli
32	247	10.0	3238	12	US-09-870-133-1	Sequence 4, Appli
33	247	10.0	3238	14	US-10-160-501-4	Sequence 42, Appli
34	247	10.0	4541	10	US-09-976-674-42	Sequence 5, Appli
35	238	9.6	3407	12	US-10-423-714-5	Sequence 6, Appli
36	238	9.6	3407	13	US-10-002-593-5	Sequence 59, Appli
37	238	9.6	3407	14	US-10-165-603-6	Sequence 63, App
38	216	8.7	2622	12	US-10-205-219-59	Sequence 22, Appli
39	214	8.6	2266	12	US-10-101-510-683	Sequence 108, App
40	214	8.6	2788	12	US-10-269-909-22	Sequence 1148, Ap
41	214	8.6	2814	10	US-09-962-832-108	Sequence 367, App
42	214	8.6	2814	10	US-09-954-456-1148	Sequence 54, Appli
43	214	8.6	2814	10	US-09-954-531-367	
44	214	8.6	2814	12	US-10-101-510-30	
45	214	8.6	2814	12	US-10-301-822-54	

#### ALIGNMENTS

RESULT 1  
US-09-976-674-8  
Sequence 8, Application US/09976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Aktinany, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Juntien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8  
LENGTH: 4523  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-8  
Alignment Scores: 1.1e-303 Length: 4523  
Pred. No.: 1

Score: 2446.50 Matches: 463  
Percent Similarity: 99.57% Conservative: 0  
Best Local Similarity: 99.57% Mismatches: 0  
Query Match: 98.57% Indels: 2  
DB: 10 Gaps: 1

US-10-070-464-5 (1-465) x US-09-976-674-8 (1-4523)

QY 1 ThrG1Thra1aenProLysVal1ThrPhelYmweSerGlu1IleMet1IleAspAlaGlu 20  
DB 1165 ACAGGTACAGCAATCTTAAGTCATCTTTAAGATGTCAGAAATTAAGATTGAGCTGAA 1224  
QY 21 GlyArlle1leAspVal1IleAspLysGluLeu1eGlnProPheGlu1IleLeuPheGlu 40  
DB 1225 GGAAGGATCATAGTGCATAGTAAGAACTAATTCACCTTTTGAGATTCTATTGAA 1284  
QY 41 GlyValGluTr1IleAlaArgAlaGlyTrpThrProGluGlyLysTrpAlaTrpSer1Ile 60  
DB 1285 GGAAGTGAATATATTTGCAAGCTGAGTGAATGCTCTAGGAAATATGCTGCTCATC 1344  
QY 61 LeuLeuAspArgSerGlnTrpArgLeuGln1IleValLeu1IleSerProGluLeuPhe1Ile 80  
DB 1345 CTACTAGATCGCTCCAGACTCGCTACAGATGTTGATCTCACTGAAATTTATTC 1404  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeu1eGluSerVal1ProAspSerVal1 100  
DB 1405 CCAAGTAGAAGATGATGTTATGAAAGGACAGACTCATTTGAGTGCCTGATTCTGTG 1464  
QY 101 ThrProLeu1IleTr1IleTrpGluGluThrThraSp1IleTrp1IleAsn1IleAsp1IlePhe 120  
DB 1465 ACCCCACTAATATCTATGAAAGAAACAACAGACTGAGTAATATCCATGACATCTTT 1524  
QY 121 HisValPheProGlnSerHisGluGluGlu1eGluPhe1IlePhe1IleAsp1IleCysAla 140  
DB 1525 CAGCTTTTCCCAAAATCAGAAAGAAATTAAGTTATTTTGCTCTGAAATGCAAA 1584  
QY 141 ThrGlyPheArgHisLeuTrpLys1IleThSer1IleLeuLysGluSerLysTrpLysArg 160  
DB 1585 ACAGGTTTCCGCTATTATACAAATTAACATCTATTAAAGAAACAAATATAAAACA 1644  
QY 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysPro1IleLysGluGlu1IleAla 180  
DB 1645 TCCAGTGGTGGGCTGCTGCTCCAGATGATTCAGAGTCCCTATCAAAAGAGAGATGCA 1704  
QY 181 IleThSerGlyGluTrpGluValLeuGlyArgHisGlySerAsn1IleGlnValAspGlu 200  
DB 1705 ATTAACGATGATGATGGAAGATCTTGCCCGCATGATCTAATATCCAAAGTTGATGAA 1764  
QY 201 ValArgArgLeuValTrpPheGluGlyThrLysAspSerProLeuGluHisLeuTrp 220  
DB 1765 GTCAGAGGCTGGTATATTTTGAAGGCAACCAAGACTCCCTTGAAGCATCACTGTAC 1824  
QY 221 ValValSerTrpValAsnProGlyGluVal1ThrArgLeuThrAspArgGlyTrpSerHis 240  
DB 1825 GTAATCGATTAAGTAATCTGAGAGAGTGCAGAGCTGACTGACCGTGGCTACTCAT 1884  
QY 241 SerCysCys1IleSerGlnHisCysAspPhePhe1IleSerLysTrpSerAsnGlnLysAsn 260  
DB 1885 TCTTGCTGCATGATGATGACACTGACTCTCTTTAAGTAAAGTATAGTAAACCAAGAT 1944  
QY 261 ProHisCysValSerLeuTrpLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1945 CCAACACTGTGTGCTCTTACAAAGCTATCAAGTCTCAAGATGACCAACCTGCAAAACA 2004  
QY 281 LysGluPheTrpAlaThr1IleLeuAspSerAlaGlyProLeuProAspTrpThrProPro 300  
DB 2005 AAGGAATTTTGGGCAACATTTTGATTCAGCAGAGTCTCTCTGACTATACTCTCCA 2064  
QY 301 GlyIlePheSerPheGluSerThrThrGlyPheThrLeuTrpGlyMetLeuTrpLysPro 320  
DB 2065 GAATTTTCTCTTTGAAAGTACTAGATTGATTTACATTTGATGAGATGCTTCAACAGCT 2124  
QY 321 HisAspLeuGlnProGlyLysTrpProThrValLeuPhe1IleTrpGlyGlyProGln 340

DB 2125 CATGATCTACAGCTCGAAAGAAATATCCTAAGTCTGCTGTTCAATATGCTGCTCG---- 2180  
QY 341 ValAla1Ile1IleAlaProVal1TrpLeuTrp1IlePheTrpAspTrpGlyTrpThGlu 360  
DB 2181 GTTGCTATGCTGGGGCCAGTCACTCTGTGATCTTCTATGATACAGGATACAGGAA 2240  
QY 361 ArgTrpMetGlyHisProAspGlnAsnGluGln1YTrpTrpLeuGlySerVal1IleMet 380  
DB 2241 CGTTATATGGGTACCTGACCGAAGTAAGAACAGGCGCATTAAGATGCTGTGGCATG 2300  
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
DB 2301 CAAGCAAGAAAGTTCCCTCTGAACCAATCGTTTACGCTTACATGATGCTTCTGAT 2360  
QY 401 GluAsnValHisPheAlaHisThrSer1IleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 2361 GAGAATGTCATTTTGACATACACAGTATATTAAGTATTTTAAAGAGGCTGGAAAG 2420  
QY 421 ProTrpAspLeuGln1IleTrpProGlnGluArgHisSer1IleArgVal1ProGluSerGly 440  
DB 2421 CCATATGATTTACAGATCTATCTCTCAAGAGACACAGCATTAAGATTCCTGAATCGGA 2480  
QY 441 GluHisTrpGluLeuHisLeuLeuHisTrpLeuGlnGluAsnLeuGlySerArg1IleAla 460  
DB 2481 GAACATATGAACTGCATCTTTTGACACTTCAAGAAACCTTGATCAGTATTCCT 2540  
QY 461 AlaLeuLysVal1Ile 465  
DB 2541 GCTCTAAAGTGATA 2555

## RESULT 2

US-10-054-776-1  
: Sequence 1, Application US/10054776  
: Publication No. US20030165818A1  
: GENERAL INFORMATION:  
: APPLICANT: Mark Robert Edbrooke  
: APPLICANT: Alan Peter Lewis  
: TITLE OF INVENTION: NOVEL PROTEIN  
: FILE REFERENCE: OG1042US  
: CURRENT APPLICATION NUMBER: US/10/054,776  
: CURRENT FILING DATE: 2002-01-23  
: NUMBER OF SEQ ID NOS: 2  
: SOFTWARE: PatentIn version 3.0  
: SEQ ID NO 1  
: LENGTH: 2649  
: TYPE: DNA  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (1)..(2649)  
US-10-054-776-1

## Alignment Scores:

Pred. No.: 6,42e-301 Length: 2649  
Score: 2422.00 Matches: 465  
Percent Similarity: 82.30% Conservative: 0  
Best Local Similarity: 82.30% Mismatches: 0  
Query Match: 97.58% Indels: 100  
DB: 12 Gaps: 1

US-10-070-464-5 (1-465) x US-10-054-776-1 (1-2649)

QY 1 ThrG1Thra1aenProLysVal1ThrPhelYmweSerGlu1IleMet1IleAspAlaGlu 20  
DB 952 ACAGGTACAGCAATCTTAAGTCATCTTTAAGATGTCAGAAATTAAGATTGAGCTGAA 1011  
QY 21 GlyArlle1leAspVal1IleAspLysGluLeu1eGlnProPheGlu1IleLeuPheGlu 40  
DB 1012 GGAAGGATCATAGTGCATAGTAAGAACTAATTCACCTTTTGAGATTCTATTGAA 1071  
QY 41 GlyValGluTr1IleAlaArgAlaGlyTrpThrProGluGlyLysTrpAlaTrpSer1Ile 60

Db 1072 GGAGTTGAATATATATGCGAGAGCTGAGTGAAGCTCTGAGGAGAAATATGCTTGCTCATC 1131  
Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
Db 1132 CTACTGATGATGCTCCAGACCTGCTGACATATGATGATCTACCTGAAATATTTATC 1191  
Qy 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
Db 1192 CCAGTGAAGATGATGTTATGAAAAGCAGAGCTACTTATGACAGTCCGTGATTCGTG 1251  
Qy 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
Db 1252 ACCCCCTAATTTATCTATGAAAGAACACACACATCGAGTAAATATCCATGACATCTT 1311  
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAspGluCysGly 140  
Db 1312 CATGTTTTTCCCAAGATCAGAAAGAAATGAGTTATTTTGTGCTCTGATGAGCAA 1371  
Qy 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
Db 1372 AAGAGTTTCCGTATTTATCAAAATTAACATCTATTTTAAAGAAAGCAAAATTAACGA 1431  
Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
Db 1432 TCCAGTGGTGGCTGCTGCTCCAGATTTCAAGTGTCTATCAAGAGAGATAGCA 1491  
Qy 181 IleThrSerGlyGluTyrProGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
Db 1492 ATTACAGATGATGAAAGGAGATTTCTTGCCGAGATGAACTATATATCCAAAGTTGATGA 1551  
Qy 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyr 220  
Db 1552 GTCAGAAAGCTGGTATATTGTAAGGACCAAGACCTCCCTTAGAGATCACTGTGC 1611  
Qy 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
Db 1612 GAGAGTACGATTAATCTTGAGAGAGGACAAAGCTGACCTGACCTACTACATC 1671  
Qy 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnIleLysAsn 260  
Db 1672 TCTTGCTGATCAGTACAGACCTGCTCTTATATAGTAAGTATGTAACCAAGAAAT 1731  
Qy 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
Db 1732 CCACACTGTGTGCTCTTACAAAGCTATCAAGCTCCTGAAGATGACCACTTGCAAAACA 1791  
Qy 281 LysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
Db 1792 AAGGAATTTGGGCCCACTTTGGATTCAGCAGGCTCTCTCTGACTATACTCTCCA 1851  
Qy 301 GluIlePheSerPheGluSerThrThrArgLysPheThrLeuTyrGlyMetLeuTyrLysPro 320  
Db 1852 GAAATTTTCTCTTTGAAAGTACTCGAATTTACCTGTATGAGAGTCTTACAAAGCT 1911  
Qy 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340  
Db 1912 CATGATCTACAGCCTGGAAGAAATCTCTACTGTCTCTCATATATGCTGCTCTG 1971  
Qy 340 ----- 340  
Db 1972 GTGAGTTGGTAATTAATCGTTTAAAGAGTCAAGTATTCGGCTTGAATACCTTAGCC 2031  
Qy 340 ----- 340  
Db 2032 TCTCTAGTTATGTGTTGTAGTATAGCAACAGGGGATCTCTGTACCGAGGCTTAA 2091  
Qy 340 ----- 340  
Db 2092 TTGGAAGGCGCTTAAATATAATAATGCGTCAAAATAGAAATTGACATCAGGTGAGGA 2151  
Qy 340 ----- 340  
Db 2152 CTCCAATATCTAGCTTCTCGATATGATTTCAATTGACTTAGATCGTGTGGCATCGACGGC 2211

Qy 340 ----- 340  
Db 2212 TGCTCTTAGAGAGATACCTCTCTCGATGCAATTAATGACAGAGTCAATATCTTCAGG 2271  
Qy 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360  
Db 2272 GTTGCTATTTGCTGGGGCCCAAGTCACTGTGATCTCTTATATATACAGATACACGGA 2331  
Qy 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
Db 2332 CGTTATATGGGTCACTCCAGCAGATGAACAGGGCTATTAATGATGATCTGTGGCATG 2391  
Qy 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
Db 2392 CAAGCAAAAAGTTCCCTCTGAAACCAATCGTTAATCTGCTTACATGATGTTCCCGAT 2451  
Qy 401 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
Db 2452 GAGAAATGCTCATTTTGCATACATACAGATATATTAATGATTTTATGAGGCTGGAAG 2511  
Qy 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGly 440  
Db 2512 CCATATGATTTACAGATCTATCTCTCAGAGAGACACAGCATAAAGATTCCTGAATCTGA 2571  
Qy 441 GluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
Db 2572 GAACATTAATGAAGCTGATCTTTTGACATCACTTCAAGAAACCTTGATCACTATTCCT 2631  
Qy 461 AlaLeuLysValIle 465  
Db 2632 GCTCTAAAGTGATA 2646  
RESULT 3  
US-10-170-789-39  
Sequence 39, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:  
APPLICANT: Rachel E. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Welch, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170,789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19

```

; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-170-789-39

Alignment Scores:
Pred. No.: 6,42e-301 Length: 2649
Score: 2422.00 Matches: 465
Percent Similarity: 82.30% Conservative: 0
Best Local Similarity: 82.30% Mismatches: 0
Query Match: 97.58% Indels: 100
DB: 12 Gaps: 1

US-10-070-464-5 (1-465) x US-10-170-789-39 (1-2649)
QY 1 ThGlyThrAlaAspProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
DB 952 ACAAGTACACCAATCTTAAGTCACTTTTAAGATGTCAGAAATATGATTTGATGCTGAA 1011
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40
DB 1012 GGAAGGATCATAGATGTCATAGATTAAGAACTAATTCACCTTTTGAGATTCATTTGAA 1071
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaATPSetrIle 60
DB 1072 GGAGTGAATATATATGCGCAAGCTGAGATGATCTCTGAGGAAATATGCTGCTCATC 1131
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
DB 1132 CTACTAATATGCTCCACGACTCGCTCAAGTATGATGTCACCTGAAATATATTTATC 1191
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
DB 1192 CCAAGTGAAGATGATGTTATGAAAGGCAAGACTCATTTGAGTCAGTCTGATTTCTGTC 1251
QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120
DB 1252 ACCCCACTAATATCTATGAAAGAAACAACAGACATCTGAGTAAATATCCATGACATCTTT 1311
QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140
DB 1312 CATGTTTTTCCCAAGTCAAGAAAGAAATGAGTTATTTTGGCTCTGATGCAAA 1371
```

```

QY 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160
DB 1372 ACAAGTTTCCGTCATTTATACAAATTTACATCTATTTTAAAGCAAAATATATAAGCA 1431
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysAspProIleLysGluIleAla 180
DB 1432 TCCAGTGTGGGCTGCTGCTCCAGAGATTTCAAGTGTCTTCAAAAGAGAGATGCA 1491
QY 181 IleThrSerGlyGluTyrPheGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200
DB 1492 ATTACCAAGTGTAAATGGAAGTTCTTGCGGCTATGATCTAATATCCAAAGTTGATGAA 1551
QY 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisIleLeuTyr 220
DB 1552 GTCAAGAGCTGTGATATTTTGAAGCACCAAGACTCCCTTTAGACATCACTTAC 1611
QY 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240
DB 1612 GTAGTCAGTTACGTAATCTGAGAGGTGACAAAGGCTGACCGTGGCTACTACAT 1671
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsn 260
DB 1672 TCTTGCTGATCAGTCAAGTCACTGTGACTTTTAAAGTAACTAATGTAACCAAGAAAT 1731
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspProThrCysLysThr 280
DB 1732 CCACACTGTGTGCTCTTTCAAGCTATCAAGTCTCAAGATATCAACCACTTGCAAAACA 1791
QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300
DB 1792 AAGGAATTTTGGGCCACCATTTTGGATTGACAGGTCCTCTCTGACTATACCTTCCA 1851
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320
DB 1852 GAAATTTTCTCTTTGAAATCTACTGATTTTAACTATGATGATGAGTCTCAACACCT 1911
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGln 340
DB 1912 CATGATCTACAGCTGTGAAGAAATATCTCATGTGTCTCATATATGATGTGCTCAG 1971
QY 340 ----- 340
DB 1972 GTGCAAGTGTGAATATCGGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTTACGC 2031
QY 340 ----- 340
DB 2032 TCTCTAGTTATGTGTTGATGATAGACACAGGGGATCTGTCAACGAGGCTTAA 2091
QY 340 ----- 340
DB 2092 TTTGAAGCGCCTTAAATATTAATGAGTCAATATGAATTAACCATCAGTGAAGAA 2151
QY 340 ----- 340
DB 2152 CTCCAATATCTACCTTCGATATGATTTCAATGACTTAGATCGTGGGCAATCCAGGC 2211
QY 340 ----- 340
DB 2212 TGGTCTTAGAGAGATACCTCTCCGTATGCAATTAATGACAGAGTCAAGTATCTTCAG 2271
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrGlu 360
DB 2272 GTTGCTATTTGCTGGGCGCCAGTCACTGTGTGATCTTCAATATACAGATCAACGAA 2331
QY 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380
DB 2332 CGTTATATGGGTCAACCTCAAGTAAACAGGGCTATTTACTTAAGATCTGTGGCATG 2391
QY 381 GlnAlaGlyLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAsp 400
DB 2392 CAAAGCAAAAGTTCCCTCTGACCAATCGTTTACTGCTTTACATGTTCTCTGGAT 2451
```

QY 401 GIUAENVAlHiePheAlahieThrSerlleLeuLeuSerPheLeuValArgAlaGlyLys 420  
DB 2452 GAGAAATGTCATTTGGACATACCACTATATTAAGTTTCTTACTGAGCGCTGAAAG 2511  
QY 421 ProTYrAspLeuGlnIleTyrrProGlnGluArgHisSerlleArgValProGlnSerGly 440  
DB 2512 CCATATGATTTTACAGATCTATCTCCTCAGAGAGACACAGCATAAAGATTCTGAATCGGGA 2571  
QY 441 GIUHeTYrGluLeuHisIleuLeuHisIleTyrrLeuGlnGluAsnLeuGlySerArglleAla 460  
DB 2572 GAACATTAAGAACTGCATCTTTGGACACTCTTCAAGAAAACTTGATCAAGTATTTGCT 2631  
QY 461 AlAlLeuLysValIle 465  
DB 2632 GCTCTAAAGTGATA 2646  
RESULT 4  
US-09-976-674-2  
; Sequence 2, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akimsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2671  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-976-674-2  
Alignment Scores:  
Pred. No.: 6,51e-301 Length: 2671  
Score: 2422.00 Matches: 465  
Percent Similarity: 82.30% Conservative: 0  
Best Local Similarity: 82.30% Mismatches: 0  
Query Match: 97.58% Indels: 100  
DB: 10 Gaps: 1  
US-10-070-464-5 (1-465) x US-09-976-674-2 (1-2671)  
QY 1 ThrGlyThrAlaAspProLysValThrPheLysMetSerGluIleuMetlleAspAlaGlu 20  
DB 959 ACAGGTACGCAAACTCTTAAGTCACTTTTAAGATGTCAGAAATATGATGATGCTGA 1018  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleuPheGlu 40  
DB 1019 GGAAGGATCATAGATGCTAGTAAAGAACTAATCAACTTTGAGATTCTAATTGAA 1078  
QY 41 GlyValGluTyrrlleAlaArgAlaGlyTrrThrProGlnGlyLysTyrrAlaTrpSerlle 60  
DB 1079 GAGATTGATATATATGCGAGACTGATGACTCTCCAGGAAAAATATCTTGATCCATC 1138  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1139 CTACTAGATCGCTCCCAAGCTGCGCTGCGAGATAGTGTATCTCACCTGAATTAATTATC 1198  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGlnSerValProAspSerVal 100  
DB 1199 CCAAGTAAAGATGATGTATGAGAAAGCAGAGACTCATGAGTCCGCTGATCTGAG 1258  
QY 101 ThrProLeuIleIleTyrrGluGluLysThrAspIleTrrlleAsnIleHisAspIlePhe 120  
DB 1259 ACGCCATTAATTTATCTATGAAAGAAACAAAGACATCTGGATAAATATCATGACATCTTT 1318

QY 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1319 CATGTTTTTCCCAAGATCAAGAGAAATTTAGTTTATTTTCCCTGATGATCAAA 1378  
QY 141 ThrGlyPheArgHisIleuTyrrLysIleThrSerlleLeuLysGlnSerIleTyrrLysArg 160  
DB 1379 ACAGGTTTCCGCATATTATACAAATATTAATCATGATTTTAAAGAAAGCAATTAACGA 1438  
QY 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
DB 1439 TCCAGTGGGGCTGCTGCTCCTCAAGTATTCMACTGCTCTATCAAGAGGAGATAGCA 1498  
QY 181 IleThrSerGlyGluTyrrGluValIleuGlyValArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1499 ATTACCAAGTGGATAGGAAGTTCTTGGCCGAGATGATCTAATATCCAAAGTTGATGAA 1558  
QY 201 ValArgArgLeuValTyrrPheGlnGlyIleTyrrLysAspSerProLeuGlnHisIleuTyrr 220  
DB 1559 GTCAAGAGGCTGGTATATTTTGAAGCACCAAGACTCCCTTTAGACATCACTGTAC 1618  
QY 221 ValValSerTyrrValAsnProGlyGluValIleThrArgLeuThrAspArgIleTyrrSerHis 240  
DB 1619 GTAGTCAGTTACGTAAATCCTGAGAGGTGACAGAGCTGACTGACCGTGTACTACAT 1678  
QY 241 SerCysValSerGlnHisCysAspPhePheIleSerLysTyrrSerAsnGlnLysAsn 260  
DB 1679 TCTTGTCATCAGTCACGACCTGACTCTTTATAGTAAAGTATGTAACGAAAGAT 1738  
QY 261 ProHisCysValSerLeuTyrrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
DB 1739 CCACACTGTGTGCTTCTTACAAAGCATCAAGCTCGAAGATGACCCACTGCAAAAC 1798  
QY 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrrThrProPro 300  
DB 1799 AAGAAATTTGGGCCACCACTTTTGATTCAGACAGGTCTCTTCTGACTATACTCTCCA 1858  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrrGlyMetLeuTyrrLysPro 320  
DB 1859 GAATTTTCTCTTTGAAAGTACTGATGATTTTACATTTATGAGAGTCTTACAAAGCCT 1918  
QY 321 HisAspLeuGlnProGlyLysLysTyrrProThrValLeuPheIleTyrrGlyGlyProGln 340  
DB 1919 CATGATCTACACCTCGAAGAAATATCTTACTGTCTCTTCTATATATGTTGTTCTCTCAG 1978  
QY 340 ----- 340  
DB 1979 GTGCAAGTGTGAATATGATTTAAAGGAGTCAAGTATTTCCGTTGAATACCTAGCC 2038  
QY 340 ----- 340  
DB 2039 TCTCTAGTTATGTTGTTGATGATAGCAACAGGGATCTGTCCACGAGGCTTAA 2098  
QY 340 ----- 340  
DB 2099 TTTGAAGGCGCTTTAAATATTAATATGTTGTTCAATATGAATATGACATGAGGAAAGA 2158  
QY 340 ----- 340  
DB 2159 CTCGAATATCTAGCTTCTCGATATGATTTCAATTGACTAGTGTGTGGCATCCAGCGC 2218  
QY 340 ----- 340  
DB 2219 TGGTCTTAGAGGATACCTCTCCCTGATGCGATTAAAGCAGAGTCAAGATCTTCAAG 2278  
QY 341 ValAlaIleAlaGlyAlaProValThrLeuTyrrIlePheTyrrAspThrGlyTyrrThrGlu 360  
DB 2279 GTTGCTATTGCTGGGGCCCAAGTCACTGTGTGATCTTCTATGATACAGGATACAGGAA 2338  
QY 361 ArgTyrrMetGlyHisProAspGlnAsnGluGlnGlyTyrrTyrrLeuGlySerValAlaMet 380  
DB 2339 CGTTATATGCGTCACCTGACCAAGAAATGAACAGGCGTATTACTTAGATCTGTGGCCATG 2398  
QY 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400

```
Db 2399 CAAGCAAGAAAAGTCCCTCGTGAACCAATCGTTACTGCTTACATGGTTCCGGAT 2458
Qy 401 GIUAARVAIHSPEALAHSETHSRSEIIEULEUSERHEULVALAGIAGLYLVS 420
Db 2459 GAGAAATGTCATTTTCCACATACCAGATATTAAGTTTAAAGAGGCTGGAAG 2518
Qy 421 PROTYASPLEUGINIEYTPROGLNUARHISERIIIEARVALPROGLUSERGLY 440
Db 2519 CCATATGATTTTACAGATCTTCTCCAGAGAGACACGCTTAAGAGTTCTGAATCGGA 2578
Qy 441 GIUHSIYIGIULEUHSIEULEUHSIYIIEUGIULUASNULEUGIYSERAGIIEAIA 460
Db 2579 GAACATATATGACATGCTTTTGCACTACCTTCAAGAAAACCTTGATCAGATTTGCT 2638
Qy 461 AIALEULYSVALILE 465
Db 2639 GCTCTAAAAGTGATA 2653

RESULT 5
US-10-170-789-37
/ Sequence 37, Application US/10170789
/ Publication No. US20030180930A1
/ GENERAL INFORMATION:
/ APPLICANT: Rachel E. Meyers
/ APPLICANT: Olandt, Peter J.
/ APPLICANT: Kapeller-Libermann, Rosana
/ APPLICANT: Curtis, Rory A. J.
/ APPLICANT: Williamson, Mark
/ APPLICANT: Welch, Nadine
/ TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
/ TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF
/ FILE REFERENCE: 10448-191001
/ CURRENT APPLICATION NUMBER: US/10/170,789
/ CURRENT FILING DATE: 2002-06-13
/ PRIOR APPLICATION NUMBER: US 09/797,039
/ PRIOR FILING DATE: 2001-02-28
/ PRIOR APPLICATION NUMBER: PCT/US01/06525
/ PRIOR FILING DATE: 2001-02-28
/ PRIOR APPLICATION NUMBER: US 60/186,061
/ PRIOR FILING DATE: 2000-02-29
/ PRIOR APPLICATION NUMBER: US 09/882,166
/ PRIOR FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: PCT/US01/19269
/ PRIOR FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 60/212,078
/ PRIOR FILING DATE: 2000-06-15
/ PRIOR APPLICATION NUMBER: US 09/934,406
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: PCT/US01/26052
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: US 60/226,740
/ PRIOR FILING DATE: 2000-08-21
/ PRIOR APPLICATION NUMBER: US 09/861,801
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: PCT/US01/16549
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: US 60/205,508
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/801,267
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07138
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,454
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/829,671
/ PRIOR FILING DATE: 2001-04-10
/ PRIOR APPLICATION NUMBER: PCT/US01/40483
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: US 60/197,508
/ PRIOR FILING DATE: 2000-04-18
/ PRIOR APPLICATION NUMBER: US 09/961,721
/ PRIOR FILING DATE: 2001-09-24
```

```
/ PRIOR APPLICATION NUMBER: PCT/US01/29904
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: US 60/235,023
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US 10/045,367
/ PRIOR FILING DATE: 2001-11-07
/ PRIOR APPLICATION NUMBER: US 60/246,561
/ PRIOR FILING DATE: 2000-11-07
/ PRIOR APPLICATION NUMBER: US 09/801,275
/ PRIOR FILING DATE: 2001-03-06
/ PRIOR APPLICATION NUMBER: PCT/US01/07074
/ PRIOR FILING DATE: 2001-03-05
/ PRIOR APPLICATION NUMBER: US 60/187,420
/ PRIOR FILING DATE: 2000-03-07
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 37
/ LENGTH: 3143
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: COS
/ LOCATION: (229)...(2874)
/ US-10-170-789-37

Alignment Scores:
Pred. No.: 8,57e-301 Length: 3143
Score: 2422.00 Matches: 465
Percent Similarity: 82.30% Conservative: 0
Best Local Similarity: 82.30% Mismatches: 0
Query Match: 97.58% Indels: 100
DB: 12 Gaps: 1

US-10-070-464-5 (1-465) x US-10-170-789-37 (1-3143)
Qy 1 ThrGlyThrAlaAsnProIysValThrPheIysMetSerGluIleMetIleAspAlaGlu 20
Db 1180 ACAGGTACAGCAAAATCTTAAGTCACTTTTAAAGATGTCAGAAATATGATGCTGGA 1239
Qy 21 GlyArgIleIleAspValIleAspIysGluLeuIleGlnProPheGluIleLeuPheGlu 40
Db 1240 GGAAGGATCATAGATTCATAGTAAGTAAGAACTTAATCAACTTTTGAATTCATTTGAA 1299
Qy 41 GlyValGluTyrIleIleAspAlaGlyTyrThrProGluGlyIleTyrAlaTyrSerIle 60
Db 1300 GGAAGTGAATATTTCCAGAGTGAATGATGATCCCTGAGGAAATATGCTGTCATC 1359
Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleIleSerProGluLeuPheIle 80
Db 1360 CTACTGATCGCTCCAGACTCCCTGCGAGATGATGTTGATCTCACCTGAATATTTATC 1419
Qy 81 ProValGluAspAspValMetGluArgGluIleGluSerValProAspSerVal 100
Db 1420 CCAATGAAAGATATTTATGAAAGCAGAGCTAATGAGTCACTGCTGATTCGAG 1479
Qy 101 ThrProLeuIleIleTyrGluGluIleThrAspIleTyrIleAsnIleHisAspIlePhe 120
Db 1480 AGCCCACTAATATATCATAGAAACACAGACATGTGATTAATTCATGACATCTTT 1539
Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIys 140
Db 1540 CATGTTTTCCTCCCAAGTACGAGAGGAAATGATTTATTTTCTCTGATGACAAA 1599
Qy 141 ThrGlyPheArgHisIleuTyrIleTyrSerIleLeuIleGluSerIleTyrIysArg 160
Db 1600 ACAGGTTTCGTCATTTATTAACAATTAATTAATTAATTAATTAATTAACGA 1659
Qy 161 SerSerGlyIleuProAlaProSerAspPheIysCysProIleIysGluIleIleAla 180
Db 1660 TCAGAGGTGGGCTGCTCTCAAGTGAATTCCTTAAGTGCCTTAAGAGGAGATGAGA 1719
Qy 181 IleThrSerGlyGluThrGluValLeuGlyArgHisIleGlySerAsnIleGlnValAspGlu 200
```



Db 1720 ATTACAGTGTGATGGAAGTTCTTGCCGGCATGATCTAATATCCAAAGTTGATGAA 1779  
 Qy 201 VALATGATGLeuVal1YrPheGluGlyThrLeuAspSerProLeuGluHisIleuLeuTyr 220  
 Db 1780 GTCCAGAGGCTGGTATATATTTTGAAGGACCAAGACCTCCCTTTAGAGCATACCTGTAC 1839  
 Qy 221 Val1AspSerTyrValAspProGluGlyVal1ThrArgLeuThrAspArgGlyTyrSerHis 240  
 Db 1840 GTAGTCAGTTACCTTAATCTGAGAGGTGACCAAGCTGACCTGCTGCTGCTACAT 1899  
 Qy 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnLeuAsn 260  
 Db 1900 TCTTGTGATCAGTCAGCAGCTGTGACTCTTATTAAGTAAAGTATAGTAAACAGAAAT 1959  
 Qy 261 ProHisCysValSerLeuTyrIleuSerSerProGluAspAspProThrCysLeuThr 280  
 Db 1960 CCACACTGTGTGCTCCCTTCAAGCTATCAAGTCTGAAATATACCACTTCCTCAAAACA 2019  
 Qy 281 LysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 300  
 Db 2020 AAGGAATTTGGGCCACATTTTGATTCAGAGCTCTTCTGACTATACCTCTCCA 2079  
 Qy 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
 Db 2080 GAAATTTTCTCTTTGAAGTACTACTGATTTACATGTATGAGATGCTCTCAAGCCT 2139  
 Qy 321 HisAspLeuGlnProGluIleuLysLysTyrProThrValLeuPheIleTyrGlyValProGln 340  
 Db 2140 CATGATCTACAGCTGGAAGAAATCTTACTGTGCTGTTCATATATGTTGTTCTCTCAG 2199  
 Qy 340 ----- 340  
 Db 2200 GTCCACTGTGTAATATCGTTTAAAGAGTCAAGATTTCCGCTTGAATACCTTAGCC 2259  
 Qy 340 ----- 340  
 Db 2260 TCTCTAGTTATGTGTGTAGTATGATGACAAACAGGGGATCTGTACCGAGGGCTTAAA 2319  
 Qy 340 ----- 340  
 Db 2320 TTTGAAGGCGCTTAAATTAATAATGGCTCAATATGAAATGACATCAGTGGAGAGA 2379  
 Qy 340 ----- 340  
 Db 2380 CTCCATATCTAGCTTCTGATATGATTTTCATGACTTAGATGTGTGGGATCCACGGC 2439  
 Qy 340 ----- 340  
 Db 2440 TGGTCTATGAGAGATACCTCTCCCTGATGACATTAATGACAGATCAGATATCTCAGG 2499  
 Qy 341 ValAlaIleAlaGlyAlaProVal1ThrLeuThrIlePheTyrAspThrGlyTyrThrGlu 360  
 Db 2500 GTTGTCTATTGTCTGGGCGCCAGTCACCTGTGTGATCTTCTATATACAGATACACGGAA 2559  
 Qy 361 ArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMet 380  
 Db 2560 CGTTATATGGGTACCTCTGACCAAGATGAACAGGGCTATTACTTAAGATTTGTGGCCATG 2619  
 Qy 381 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 400  
 Db 2620 CAAGCAGAAAAGTTCCCTCTGAACCAATCGTTTACTGCTCTTACATGGTTTCCGGAT 2679  
 Qy 401 GluAsnValHisPheHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 420  
 Db 2680 GAGAAATGTCATTTTCACTACATACAGATATTAATCAAGTTTATATGAGGCTGGAAGA 2739  
 Qy 421 ProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgVal1ProGluSerGly 440  
 Db 2740 CCATATGATTTTACAGATCTATCTCTCAGAGAGACACAGCATTAAGATTTCTGATCGGA 2799  
 Qy 441 GluHisTyrGluLeuHisIleLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAla 460  
 Db 2800 GAACATTAAGAACTGCATCTTTTGGACATACCTTCAAGAAAACCTTGATACAGTATTGCT 2859

Qy 461 AlaLeuLeuVal1Ile 465  
 Db 2860 GCTCTAAAAGTGTATA 2874  
 RESULT 6  
 US-09-976-674-20  
 ; Sequence 20, Application US/09976674  
 ; Patent No. US20020115843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qi, Steve  
 ; APPLICANT: Akinsanya, Karen  
 ; APPLICANT: Riviere, Pierre  
 ; APPLICANT: Junien, Jean-Louis  
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTV  
 ; FILE REFERENCE: 70669  
 ; CURRENT APPLICATION NUMBER: US/09/976, 674  
 ; PRIOR FILING DATE: 2001-10-12  
 ; PRIOR FILING DATE: 2000-10-12  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 20  
 ; LENGTH: 4676  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-976-674-20  
 Alignment Scores:  
 Pred. No.: 4,74e-300 Length: 4676  
 Score: 2418.50 Matches: 463  
 Percent Similarity: 90.10% Conservative: 1  
 Best Local Similarity: 89.90% Mismatches: 1  
 Query Match: 97.44% Indels: 51  
 DB: 10 Gaps: 1  
 US-10-070-464-5 (1-465) x US-09-976-674-20 (1-4676)  
 Qy 1 ThrGlyThrAlaAsnProLysVal1ThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
 Db 1165 ACAGGTACAGCAAAATCTTAAGTCACTTTAAGATGTCAGAAATATGATGTGCTGAA 1224  
 Qy 21 GlyArgIleIleAspVal1IleAspLysGluLeuIleGluProPheGluIleLeuPheGlu 40  
 Db 1225 GGAAGATCATATGATTCATATGATTAAGAACTAATTCACCTTTGAGATCTAATTTGAA 1284  
 Qy 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 Db 1285 GAGGTGAATATATATCCAGAGCTGATGACTCCAGAGGAAATATGCTGTGCATC 1344  
 Qy 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 Db 1345 CTACTAGATCGCTCCAGACTGCGCTTACAGATGTGTGATCTCACTGAATATTATTC 1404  
 Qy 81 ProValGluAspAspVal1MetGluArgGluIleGluSerVal1ProAspSerVal 100  
 Db 1405 CCAGTGAAGATGATGTATGAAAGCAGAGCTATTGAGTCACTGCTGATTCGTG 1464  
 Qy 101 ThrProLeuIleIleTyrGluGluThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 Db 1465 ACCGCTAATATATCTATATGAAGAAACACAGACATCTGATTAATATCCATGACATCTT 1524  
 Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLys 140  
 Db 1525 CATGTTTTTCCCAAGTACACAGAGGAAATGAGTTTATTTTCCCTCAATACCAA 1584  
 Qy 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 Db 1585 ACAGGTTTCGTCATTTATTAACAAATATACATCTATTTTAAAGGAAAGCAAAATTAACGA 1644  
 Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 Db 1645 TCCAGTGTGGGCTGCTGCTCCAAAGTATTTCAAGTGTCTATCAAAAGAGGATAGCA 1704

```

QY 181 ILeThSerGIyGluTrpGluValIleuGIyARhIseGIySerAsnIleGIuValaAsGlu 200
DB 1705 ATTACCAAGTGGTGAATGGAGAGTTCTGGCCGGCAGATGATTAATCCAAAGTTGATGAA 1764
QY 201 ValArgArgLeuValIyTrpGluGIyThrLysAspSerProLeuGIuIshIseLysTy 220
DB 1765 GTCAGAAAGCTGGTATATTGTAAGGACCAAGAGTCCCTTTAGAGCATCACTGTAC 1824
QY 221 ValValSerTyTrpAlaAsnProGIyGluValThrArgLeuThrAspArgGIyTrpSerHis 240
DB 1825 GTAGTACAGTACCTAAATCTGGAGAGGATCAAGGCTGACCTGCTGCTACAT 1884
QY 241 SerCySeGIyIleSerGIuIshIseGIyAspPheIleSerTyTrpSerAsnGIuLysAsn 260
DB 1885 TCTTGCTGCATCACTGCACCTGTGACTTTTATAGTAAGTATAGTAAACCAAGAAAT 1944
QY 261 ProHisCyValSerLeuTyTrpLysLeuSerSerProGIuAspAspProThrCyValSer 280
DB 1945 CCACACGTGTGTCCTTTACAAAGCTATCAAGCTCTGAAAGATGACCCAACTTCGCAAAACA 2004
QY 281 LysGluPheTrpAlaThrIleLeuAspSerAlaGIyProLeuProAspTyTrpProPro 300
DB 2005 AAGGAATTTGGCCACCATTTGGATTGACAGAGTCTCTCTGACTAACTCTCTCA 2064
QY 301 GluIlePheSerPheGluSerThrThrGIyPheThrLeuTyTrpLysLeuTyTrpLysPro 320
DB 2065 GAATTTTCTCTTTGAAATACACTGATTTTACATGATTTATGGAGTCTCTCAAGACT 2124
QY 321 HisAspLeuGIuIshIseGIyLysTyTrpProThrValLeuPheIleTyTrpGIyPro 339
DB 2125 CATGATCTACAGCTGGAAAGAAATATCTCATGTGCTTCAATATGTGTGTGCG-GGT 2183
QY 339 ----- 339
DB 2184 CAAATGAAATTCAGATCCAGTCAGTGAAGAGCTCCAAATATCTAGCTTCTCATATGATTC 2243
QY 339 ----- 339
DB 2244 ATTGACTTACATGCTGTGGCATCCAGCTGGTCTCATGAGAGATACCTCTCCCTGATG 2303
QY 340 -----GlnValAlaIleAlaGIyAlaProValThrLeu 350
DB 2304 GCATTAATGACAGAGTCAGATATCTTCAGGGTGTCTATGCTGGGGCCCAAGTCACCTG 2363
QY 351 TrpIlePheTyAspThrGIyTrpThrGIyArgTyTrpMetGIyHisProAspGIuAsnGIu 370
DB 2364 TGGATCTTCTATATACAGATACAGAAAGTTATATGGGTCACTGACCAAGATGAA 2423
QY 371 GlnGIyTyTrpLeuGIySerValAlaMetGlnAlaGIuLysPheProSerGIuProAsn 390
DB 2424 CAGGGCTATTAAGGATCTGTGGCCATGCAAGCAAAAGTTCCCTCTGAACCAAT 2483
QY 391 ArgLeuLeuLeuLeuHisGIyPheLeuAspGIuAsnValHisPheAlaHisTrpSerIle 410
DB 2484 CGTTTACTGCTTACATGATGTTCCTGAGTGAAGATTCATTTTGGACATACACAGATA 2543
QY 411 LeuLeuSerPheLeuValArgAlaGIyLysProTyAspLeuGIuIleTyTrpProGIu 430
DB 2544 TTTACTAGTTTATAGAGAGGCTGGAAGCCATATATATTAAGATCTATCTCTCAGAG 2603
QY 431 ArgHisSerIleArgValIProGIuSerGIyGluHisTyTrpGIuLeuHisLeuLeuHisTy 450
DB 2604 AGACACAGCATTAAGATTCTCGAATCGGAGAACATTAATACATGATCTTTTGGACTAC 2663
QY 451 LeuGIuGIuAsnLeuGIySerArgIleAlaIleLeuLysValIle 465
DB 2664 CTTCAAGAAACCTTGGATCACTATTGCTCTTAAAGTGATA 2708

```

RESULT 7  
 US-09-976-674-12  
 ; Sequence 12, Application US/0976674  
 ; Patent No. US20020115843A1

```

; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akimane, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 12
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-976-674-12

Alignment Scores:
Pred. No.: 6,7e-298 Length: 4829
Score: 2402.00 Matches: 465
Percent Similarity: 82.01% Conservative: 0
Best Local Similarity: 82.01% Mismatches: 0
Query Match: 96.78% Indels: 102
DB: 10 Gaps: 1

US-10-070-464-5 (1-465) x US-09-976-674-12 (1-4829)

QY 1 ThrGIyThrAlaAsnProLysValIlePheLysMetSerGIuIleMetIleAspAlaGIu 20
DB 1165 ACAGTACAGCAAAATCTTAAGTCACTTTAAGATGTCAGAAATATGATTTGATGCGAA 1224
QY 21 GlyArgIleIleAspValIleAspLysGIuLeuIleGIuProGluIleLeuPheGlu 40
DB 1225 GGAAGATCATATATCTCATATGATTAAGAACTTAATCACTTTGATGATTTGAA 1284
QY 41 GlyValGIuTyTrpIleAlaArgAlaGIyTrpThrProGIuGIyLysTyTrpAlaTrpSerIle 60
DB 1285 GGAGTTGATATATTTCCAGAGCTGATGATGATCTCTGAGGAAATATGTTGCTGATC 1344
QY 61 LeuLeuAspArgSerGIuThrArgLeuGIuIleValLeuIleSerProGIuLeuPheIle 80
DB 1345 CTACTGATCGCTCCAGACTCCCTACAGATGATGTTGATCTCACCTGATTAATTTATC 1404
QY 81 ProValGIuAspAspValMetGIuArgGIuArgLeuIleGIuSerValProAspSerVal 100
DB 1405 CCAGTGAAGATGATGTATGGAAGGACAGACTGATTAAGTCAAGTCTGATTCGTG 1464
QY 101 ThrProLeuIleIleTyTrpGIuGIuThrThrAspIleTrpIleAsnIleHisAspIlePhe 120
DB 1465 AGCCACTAATTAATCTATGAGAAACACAGACATCTGATTAATATTCATGACATCTTT 1524
QY 121 HisValPheProGIuSerHisGIyGIuGIuIleGlnPheIlePheAlaSerGIuCyLys 140
DB 1525 CATGTTTTTCCCAAGTACAGAAAGAAATGAGTTTATTTTCTCTGAAATCCAA 1584
QY 141 ThrGIyPheArgHisLeuTyTrpLysIleThrSerIleLeuLysGIuSerTyTrpLysArg 160
DB 1585 ACAGTTTCCGTCATTTATTAACAAATTAATCACTATTTTAAAGGAAACCAATATTAACGA 1644
QY 161 SerSerGIyGIuLeuProAlaProSerAspPheLysCyAspProIleLysGIuGIuIleAla 180
DB 1645 TCCAGTGGTGGGCTGCTCTCCAAATGATTTCAAGTGTCTTATCAAAAGAGGATAGCA 1704
QY 181 ILeThSerGIyGluTrpGluValIleuGIyARhIseGIySerAsnIleGIuValaAsGlu 200
DB 1705 ATTACCAAGTGGTGAATGGAGAGTTCTGGCCGGCAGATGATTAATCCAAAGTTGATGAA 1764
QY 201 ValArgArgLeuValIyTrpGluGIyThrLysAspSerProLeuGIuIshIseLysTy 220
DB 1765 GTCAGAAAGCTGGTATATTGTAAGGACCAAGAGTCCCTTTAGAGCATCACTGTAC 1824

```

QY 221 ValValSerTyrValAsnProGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
DB 1825 GTAGTCACTACGTAACAAATCTCGAGAGAGTGAACAAGCTGACTGACCTGCTACTCAT 1684  
QY 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerIleTyrSerAsnGlnLysAsn 260  
DB 1885 TCTTGCTGCATCAGTCACTGACCTGCTCTTTAATAGTAAGTAACTAACCAGAGAAAT 1944  
QY 261 ProHisCysValSerLeuTyrLysLeuSerSerProLysAspPheProTyrCysLysThr 280  
DB 1945 CCACACGTGTGTCCTTACAAAGCTATCAAGTCTCGAAGATACCCAACTTCGAAACA 2004  
QY 281 LysGluPheIlePheIleThrIleLeuAspSerIleGlyProLeuProAspTyrThrProPro 300  
DB 2005 AAGCAATTTGGGCGACCACTTTGGATTGACAGAGTCTCTTCTGACTATCTCTCCCA 2064  
QY 301 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 320  
DB 2065 GAAATTTTCTCTTTGAAAGTACTACTGATTTACATTTATGAGGATGCTCTCAAGCCT 2124  
QY 321 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--ProG 340  
DB 2125 CATGATCTACAGCTGGAAGAAATATCTTACTGTGCTGTTCATATATGAGTGTCTCTC 2184  
QY 340 1n----- 340  
DB 2185 AGGTGCAAGTTGCGAATTAATCGGTTTAAAGAGTCAAGTATTTCCGTTGAATACCTAG 2244  
QY 340 ----- 340  
DB 2245 CCTCTAGGTATGTGTGTATGATATAGCAACAAGGAGTCTCTGACCGAGGCTTA 2304  
QY 340 ----- 340  
DB 2305 AATTGGAAGCGCCTTAAATATTAATGAATGGGTCAAAATGAAATTGACATCAGGTGAAG 2364  
QY 340 ----- 340  
DB 2365 GACTCAATATCTAGCTTCTCGATATGATTTTCACTTACCTAGATCGTGGGATCAGCAG 2424  
QY 340 ----- 340  
DB 2425 GCTGATCTATGAGAGATACCTCTCCCTGATGCAATTAAATGACAGAGTCAATATCTTCA 2484  
QY 341 --ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThrG 360  
DB 2485 GGGTGTCTATGCTGGGCCCCAGTCACTGTGGATCTTATGATACAGGATACAGCAG 2544  
QY 360 LysArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAla 380  
DB 2545 AACGTTATATGGGTCACTGACCAAGAAATGAACAGGGCTATTAAGATCTGTGGCA 2604  
QY 380 eTGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 400  
DB 2605 TCACAAGCAAAAAAGTCCCTCTGAAACCAAAATGTTACTGCTTCAACAGGTTCTCTG 2664  
QY 400 ArgGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlyAla 420  
DB 2665 ATGAGAGATTCATTTTGCACATACCAATATATTAAGTGTGTTTAAAGAGGCGCGAA 2724  
QY 420 YSPProTyrAspLeuGlnIleTyrProGlnGlnLysArgHisSerIleArgValProGluSerG 440  
DB 2725 AGCCATATGATTTAAGATCTATCTCCAGAGAGAGACAGCATMAAGTCTCTGAATCGG 2784  
QY 440 LysGluHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGlnLysSerArgIleAla 460  
DB 2785 GGAACATTTATGAACGATCTTTTGGCACTACCTTCAAGAAAACTTGATCACGATATG 2844  
QY 460 1aAlaLeuLysValIle 465  
DB 2845 CTGCTCTAAAAGTGATA 2861  
RESULT 8

US-09-976-674-22  
; Sequence 22, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
; FILE REFERENCE: 70669  
; CURRENT APPLICATION NUMBER: US/09/976,674  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,117  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 4685  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-976-674-22  
Alignment Scores:  
Pred. No.: 2,6e-265 Length: 4685  
Score: 2149.00 Matches: 428  
Percent Similarity: 83.65% Conservative: 7  
Best Local Similarity: 82.31% Mismatches: 28  
Query Match: 86.58% Indels: 58  
DB: 10 Gaps: 6  
US-10-070-464-5 (1-465) x US-09-976-674-22 (1-4685)  
QY 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20  
DB 1165 ACAGGTACAGCAAAATCTTAAGTCACTTTAAGATGTCAGAAATATGATGATGCTGAA 1224  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1225 GGAAGGATCATAGATGTCTATGATTAAGAACTTAATCACTTTTGAATTCATTTTGA 1284  
QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
DB 1285 GGAAGTGAATATATATGTCAGAGCTGATGATCTCCGAGGAAATATGCTTGCTCATC 1344  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1345 CTACTAGATCGCTCCAGACTCGCTTACAGATAGTGTATGATCTACCTGAATATTATTC 1404  
QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1405 CCAAGTGAAGATGATGTTATGGAAGGACAGACATCTTGAATGAGTCACTGCTGTCG 1464  
QY 101 ThrProLeuIleIleTyrGluGlnLysThrAspIleTyrIleAsnIleHisAspIlePhe 120  
DB 1465 AGCCCATATATATCTATGAAAGAAACACAGACATCTGATTAATATTCATGACATCTTT 1524  
QY 121 HisValPheProGlnSerHisGluGlnGluIleGluPheIlePheAlaSerGluCysLys 140  
DB 1525 CATGTTTTCCTCCCAAGTACAGAGAGGAAATGAGTTTATTTTGGCTCTGAATCGAA 1584  
QY 141 ThrGlyPheArgHisLeuTyrIleThrSerIleLeuLysGluSerIleTyrLysArg 160  
DB 1585 ACAGGTTTCCTGATTTATACAAATTTTCATCTATTTTAAAGAAAGCAAAATATTAACGA 1644  
QY 161 SerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluGlnIleAla 180  
DB 1645 TCCATGTGTGGGCTCTCTCCCAAGTATTTCAAGTGTCTCTATCAAGAGAGATAGCA 1704  
QY 181 IleThrSerGlyLysTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAspGlu 200  
DB 1705 ATTACAGGTGATGATGGAAGTTCTTGCGCGCATGATCTTAATATCAAGATTGATGA 1764  
QY 201 ValArgArgLeuValTyrPheGluGlnLysThrLysAspSerProLeuGlnHisIleLeuTyr 220

```

Db      1765 GTCAGAGGCTGGTATATTGTAAGGACCAAGACCTCCCTTAAAGCATCAGCTGAC 1824
QY      221 ValValSerTyrValAsnProGlyGlyValThrArgLeuThrAspArgGlyTyrSerHis 240
Db      1825 GTAAGTAGTACCTAAATCTGGAGAGGTACCAAGGCTGACCGTGGCTACTACAT 1884
QY      241 SerCysCysIleSerGlnHisCysAspPhePheIleSerTyrSerAsnGlnLysAsn 260
Db      1885 TCTGCTGCATCAGTACGACCTGACTCTTATATAGTATAGTATACCAAGAAAT 1944
QY      261 ProHisCysValSerLeuTyrLysLeuSerSerProGlnAspAspProThrCysLysThr 280
Db      1945 CCACACTGCTGTCCCTTACCAAGCTATCAAGCTCAAGATGATGCCAACCTTCAAAACA 2004
QY      281 LysGlnPheThrPheIleThrLeuAspSerAla----- 291
Db      2005 AAGGAATTTTGGCCACCAATTTGGATTCAGTCTCCTCAGGTGACGCTGGATTAATCGGT 2064
QY      292 -----GlyProLeuProAspTyrThrProPro 300
Db      2065 TTAAGAGTCAGATATTCCGCTGATATACCTAGCTCTCTA--GGTTATGTGCTTGA 2123
QY      301 GluIle-----PheSerPheGluSerThrThrGlyPhe 311
Db      2124 TGATATGACACACAGGGGATCTGTCAACGAGGCTTAAATTTGAAGCGCTTAAATAT 2183
QY      312 ThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThr 331
Db      2184 AAAATG--GGTCAATAA--GAAATTGACATCAGGTGAGAGGACTCCAAATATCTAGCT 2237
QY      332 ValLeuPhe-----IleTyrGlyGly 338
Db      2238 TCTCGATATGATTCTTATGACTTATAGATTCGTGTGGGATCCAGGCTGTGCTTATGAAGA 2297
QY      339 Pro-----GlnValAlaIleAlaGly 345
Db      2298 TACCTCTCCCTGATGCAATTATGACAGAGTCAGATATCTTCAGGGTGTCTATGTCTGG 2357
QY      346 AlaProValThrLeuThrIlePheTyrAspThrGlyTyrThrGlnArgTyrMetGlyHis 365
Db      2358 GCCCGCATCTCTGTGATCTTCTATGATACAGATACCGGAAGCTTATATGGGCTAC 2417
QY      366 ProAspGlnAsnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyLysPhe 385
Db      2418 CCGTACACAGATTAACAGGCTATTCTTAGATCTGTGGCCATGCAAGAGAAAAGTTC 2477
QY      386 ProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGlnAsnValHisPhe 405
Db      2478 CCTCTGAACCAATAGTTTACTGCTCTTACATGATGTTTCTGTGATAGAAATGCCATTTT 2537
QY      406 AlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGln 425
Db      2538 GACATACCAAGTATATTAAGATTTTATAGTGGGCTGGAAAGCCATATGATTTACAG 2597
QY      426 IleTyrProGlnGlnArgHisSerIleArgValProGlnSerGlyGlnHisTyrGlnLeu 445
Db      2598 ACTTATCTCAGAGAGACACAGCATTAAGATTCCTGAATCCGGAGAACATTATGAACGTG 2657
QY      446 HisLeuLeuHisTyrLeuGlnGlnLysLeuGlySerArgIleAlaIleLeuLysValIle 465
Db      2658 CATCTTTTCACTTCACTTCAAGAAAACCTTGATCACTGATGCTCTCTTAAAGATATA 2717

```

```

; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2617
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-976-674-4

Alignment Scores:
Pred. No.: 2,24e-181 Length: 2617
Score: 1494.50 Matches: 261
Percent Similarity: 62.958 Conservative: 69
Best Local Similarity: 50.548 Mismatches: 105
Query Match: 60.21% Indels: 101
DB: 10 Gaps: 2

US-10-070-464-5 (1-465) x US-09-976-674-4 (1-2617)

QY      1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
Db      932 ACAGGACAGCAAGAAATCCCAAGATTGCTTGAATGCTGAGATTCCAGACTGCACAGCAG 991
QY      21 GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGlnIleLeuPheGlu 40
Db      992 GGCAGATATGCTTCGACCCAGCAAGAGAGCTGTGTACACCTTCACGCTCGCTGTTCCCG 1051
QY      41 GlyValGlnTyrIleAlaArgAlaGlyTyrThrProGlnGlyLysTyrAlaTyrSerIle 60
Db      1052 AAGGTGAGTATCATCCCGACCGGCTGTGACCCGGAGTGGCAAAATACCTGGCCATG 1111
QY      61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
Db      1112 TTCCTGACCGGCCCCCAGCAGTGTCTCCAGCTCTCTCCCGCCGCGCTGTTCATC 1171
QY      81 ProValGlnAspAspValMetGlnArgGlnArgLeuIleGluSerValProAspSerVal 100
Db      1172 CCGAGCACAGAAATGAGAGAGCAAGCTAGCTCTGCCAGAGACTGTCCCAAGAAATGTC 1231
QY      101 ThrProLeuIleIleTyrGlnGlnLysThrAspIleTyrIleAsnIleHisAspIlePhe 120
Db      1232 CAGCCTATGTGTGTACAGAGAGTCAACCAACGTCTGATCAATGTTTATGATCATCTTC 1291
QY      121 HisValPheProGlnSerHis--GlnGlnGlnIleGluPheIlePheAlaSerGlnCys 139
Db      1292 TATCCCTCCCAATCAAGAGGAGAGCAGAGCTCTCTCTCCGCGCAATGAATGC 1351
QY      140 LysThrGlnPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159
Db      1352 AAGACCGGCTTCTCCCATTTGTAAGATCAACCGCGCTTTTAAATCCAGAGGCTAGCAT 1411
QY      160 ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGlnGlnIle 179
Db      1412 TGAAGTGAACCTTCAGCCCGGGAAGATGAATTTAACTGATGATTAAGAGAGATTT 1471
QY      180 AlaIleThrSerGlyGlnTyrProLysValLeuGlyArgHisGlySerAsnIleGlnValAsp 199
Db      1472 GCTTGACCAAGGAGTGAATGGAGGTGTTGGAGAGCCAGGCTCCAAAGATCTGGGCTCAAT 1531
QY      200 GlyValArgArgLeuValTyrPheGlnGlnTyrLysAspSerProLeuGlnHisIleAsn 219
Db      1532 GAGGAGACCAAGCTGTGATCTTCCAGGACCAAGACACCGCTGAGACCAACCTTC 1591
QY      220 TyrValIleSerTyrValAsnProGlnGlyGlnValThrArgLeuThrAspArgGlyTyrSer 239
Db      1592 TACGTGTGATGAGAGCGGCGGCGAGATCGATACGCTTCAACCGCGGCTTCTCC 1651
QY      240 HisSerCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259

```

RESULT 9  
US-09-976-674-4

; Sequence 4, Application US/09976674  
; Patent No. US20020115843A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Steve  
; APPLICANT: Akinsanya, Karen  
; APPLICANT: Riviere, Pierre  
; APPLICANT: Junien, Jean-Louis  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPP1V

```
Db 1652 CATAGTGTCTCCATGAGCCGAACTTCGACATGTCCTCAGCCACTAGACAGCGTAGC 1711
QY 260 AapProHisCysValSerLeuTyrLysLeuSerSerProGluAapProThrCysLys 279
Db 1712 AGCGCCCGCTGCGTGCAGCTCTACAGTACAGCGCGCCGCGACGACCCCTGCGACAG 1771
QY 280 ThrLysGluPheThrPalaThrIleLeuAapSerAlaGlyProLeuAapTyrThrPro 299
Db 1772 CAGCCCGCGTCTTGGGCTAGCATGATGAGAGCGACGCTGCCCGGATTAATGTTCTT 1831
QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 319
Db 1832 CCAGAGATCTTCATCTTCCACAGCGCTCGATGTCGCGCTTACGCGCATGATCTACAAG 1891
QY 320 ProHisAapLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLysPro 339
Db 1892 CCCCAGCCCTTGGACGCGGAGGAGAGACCCACCGCTCTTGTGATATGAGAGGCCCC 1951
QY 340 Gln----- 340
Db 1952 CAGGTGCAGTGTGATATACCTCTTCAAGGACATCACTTGGGGCTCAACACACTG 2011
QY 340 ----- 340
Db 2012 GCCTCCCTGGGCTACGCGCTGTGTGATGACGCGAGGGCTCTGTGACGAGGGCTT 2071
QY 340 ----- 340
Db 2072 CGGTTGAAAGGGGCGCTGAAACCAAAATGGGCGAGGTGAGATCGAGACAGGTGAG 2131
QY 340 ----- 340
Db 2132 GGCCTGACAGTGTGCGCGAGAGATAGGCTTATGACCTGAGCGGAGTGCATCAT 2191
QY 340 ----- 340
Db 2192 GCGTGTCTTACGAGGGGCTTCTCTGCTCATGAGGGCTAATCCAGAGCCGAGGTTC 2251
QY 341 ---ValAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 359
Db 2252 AAGGTGGCATGCGCGGAGTCCCGGATCCGCTGATGAGTGGCTTCCACACAGAGGTACT 2311
QY 360 GluArgTyrMetGlyHisProAapGlnAangLugInGlyTyrTyrLeuGlySerValAla 379
Db 2312 GAGCGCTACATGAGAGCTCCCTGAGAACACACAGCAGGCTATGAGCGGCTTCGAGGC 2371
QY 380 MetGlnAlaGluLysPheProSerGluProAapArgLeuLeuLeuHisGlyPheLeu 399
Db 2372 CTGCAAGTGGAGAGAGCTGCCCAATGAGCCCAACCGCTTCTTATCTCCACAGGCTTC 2431
QY 400 AapGluAapValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419
Db 2432 CAGGAAAGCTGACCTTTTCCACACAAACCTTCTGCTCCCAATGATTCGAGAGAGG 2491
QY 420 LysProTyrAapLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
Db 2492 AAACCTTACCACTCCAGATCTACACCCCAAGAGAGACACAGTATTCGCTCCCGAGTGC 2551
QY 440 GlyLysHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAapLeu 455
Db 2552 GCGGAGCACTATGAGAGTCAAGTGTGCTGCACTTCTACAGGAATATACCTC 2599
-
RESULT 10
US-09-976-674-28
; Sequence 28, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
; FILE REFERENCE: 70669
```

```
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 4219
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-28
Alignment Scores:
Pred. No.: 5,01e-181 Length: 4219
Score: 1494.50 Matches: 281
Percent Similarity: 62.95% Conservative: 69
Best Local Similarity: 50.54% Mismatches: 105
Query Match: 60.21% Indels: 101
DB: 10 Gaps: 2
US-10-070-464-5 (1-465) x US-09-976-674-28 (1-4219)
QY 1 ThrGlyThrAlaAapProLysValThrPheLysMetSerGluLeuMetIleAapAlaGlu 20
Db 1288 ACAGGAGACAGAGATCCCAAGATTGCTTGAATACTCGTAGTTCAGACTGACAGCCAG 1347
QY 21 GlyArgIleIleAapValIleAapLysGluLeuIleGlnProPheGluIleLeuPheGlu 40
Db 1348 GGCAGAGATGCTCGACCCAGAGAGAGAGAGTGTGTCAGACCTTACGCTGCTGCTCCG 1407
QY 41 GlyValGluTyrIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 60
Db 1408 AAGGTGAGTACATGCGCGGCGCGGTGACCCGAGATGGCAAAATACCTGCGCGCATG 1467
QY 61 LeuLeuAapArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
Db 1468 TTCCTGAGACCGGCGCCAGAGAGTGTCCAGCTGCTCTCTCTCCCGGCGCTGTTCTATC 1527
QY 81 ProValGluAapAapValMetGluArgGlnArgLeuIleGluSerValProAapSerVal 100
Db 1528 CCGAGCACAGAAATGAGAGAGCGGCTAGCTCTGCGAGAGCTGTCGCCAGAGATGTC 1587
QY 101 ThrProLeuIleIleTyrGlnGluLysThrAapIleThrIleAapIleHisAapIlePhe 120
Db 1588 CAGCGGTATGTGTGACAGAGAGGTACCAACGCTGTGATCAATGTTTATGATACATCTTC 1647
QY 121 HisValPheProGlnSerHis---GluGlnGluIleGluPheIlePheAlaSerGluCys 139
Db 1648 TATCCCTTCCCAATCAGAGAGAGAGAGAGAGCTCTGCTTCTCGCGCCAAATGATTC 1707
QY 140 LysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159
Db 1708 AAGACCGGCTTGTGCATTTGTACAAAGTCAACCGCGTTTAAATCCACAGGGCTACGAT 1767
QY 160 ArgSerSerGlyGlyLeuProAlaProSerAapPheLysCysProIleLysGluGluIle 179
Db 1768 TGGAGTGAAGCCCTTACGCCCGGAGAGATGAATTAAGTGCCTTAAAGAGAGATT 1827
QY 180 AlaIleThrSerGlyGluTyrProGluValLeuGlyArgHisGlySerAapIleGlnValAap 199
Db 1828 GCTGACACAGCGCGTAATGGAGAGTTCGCGAGAGACAGGCTCCAAAGTTCGGGTCAAT 1887
QY 200 GluValArgArgLeuValTyrPheGluGlyThrLysAapSerProLeuGluHisLeu 219
Db 1888 GAGGAGACCAAGCTGTGTACTTCCAGAGGACCAAGAGACAGCGCTGAGAGACCAACCTC 1947
QY 220 TyrValIleSerTyrValAapProGlyGluValThrArgLeuThrAapArgGlyTyrSer 239
Db 1948 TACGTGTCAGCTATGAGAGCGCGCGGAGATCTTACCGCTTACACACCGCCGCTTCTCC 2007
QY 240 HisSerCysHisSerGlnHisCysAapPhePheIleSerLysTyrSerAandLys 259
Db 2008 CATAGCTCTCCATGAGACCAACTTGCATGTTCTGACGCCACTACAGCAGCGTGAAC 2067
```













QY 375 euglyserValAlaMetGlnAlaGluYrPheProSerGluProAsnArgLeuLeu 395  
DB 2070 TAGATCTGTGGCATGCAAGCAAGAAAGTTCCTTGAACCAATCGTTACTGCTC 2129  
QY 395 euHlsglyPheLeuAspGluAsnValHlPheAlaHlStrSerHlLeuLeuSerPhe 415  
DB 2130 TACATGGTTTCTTGATGAAATGTCATTTTGCATACAGATATATCTAGATTTT 2189  
QY 415 euValArgAlaGlyrProTyAspLeuGlnLeuTyProGlnGluArgHlSerLea 435  
DB 2190 TAGTGAGGCTGGAAGCCATATGATTTACAGATCTATCTCAGAGAGACAGCATAA 2249  
QY 435 rglValProGluSerGlyGluHlStrGlyLeuLeuHlStrLeuGlnGluAsn 455  
DB 2250 GAGTCTCTGATTCGGGAGAACATTTATGACTGCATCTTTTGCATCTACCTTCAAGAAAC 2309  
QY 455 euglySerArgLeuAlaAlaLeuValLeu 465  
DB 2310 TTGATCAGCTATTCGCTGCTTAAAGTGATA 2341

RESULT 15  
US-09-976-674-32  
Sequence 32, Application US/09976674  
Patent No. US20020115843A1  
GENERAL INFORMATION:  
APPLICANT: Qi, Steve  
APPLICANT: Akinsanya, Karen  
APPLICANT: Riviere, Pierre  
APPLICANT: Junien, Jean-Louis  
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV  
FILE REFERENCE: 70669  
CURRENT APPLICATION NUMBER: US/09/976,674  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,117  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 32  
LENGTH: 4076  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-976-674-32

Alignment Scores:  
Pred. No.: 1,856-161 Length: 4076  
Score: 1342.50 Matches: 259  
Percent Similarity: 64.83% Conservative: 71  
Best Local Similarity: 50.88% Mismatches: 125  
Query Match: 54.09% Indels: 54  
DB: 10 Gaps: 4

US-10-070-464-5 (1-465) x US-09-976-674-32 (1-4076)

QY 1 ThrGlyThrAlaAsnProIysValThrPheLysMetSerGluLeuMetIleAspAlaGlu 20  
DB 1288 ACGGAGCAAGAAATCCCAAGATTCCTTGAACCTGCTGATTCACAGACTGCAGCGAG 1347  
QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40  
DB 1348 GCGAAGTCTCTCGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407  
QY 41 GlyValGluTyrlleAlaArgAlaGlyTyrlleProGluGlyLysTyrlleAlaTrpSerIle 60  
DB 1408 AAGGTGAGATACATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1467  
QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
DB 1468 TTCCTGACCGGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527  
QY 81 ProValGluAspArgValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
DB 1528 CCGAGACAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1587

QY 101 ThrProLeuIleIleTyrlleGluGluThrThrAspIleTrpIleAsnIleHlAspIlePhe 120  
DB 1588 CAGCCCTATGTGTGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1647  
QY 121 HlValPheProGlnSerHlStrGluGluIleGluPheIlePheAlaSerGluCys 139  
DB 1648 TATCCCTTCCCAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1707  
QY 140 TyrlleGlyPheArgHlStrLeuTyrlleHlStrSerIleLeuIleGluSerGlyTyrlle 159  
DB 1708 AAGACCGGCTTCGACATTTGTAACAAGTACCGCGTTTAAATCCAGAGGCTACGAT 1767  
QY 160 ArgSerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLysGluIle 179  
DB 1768 TGGAGTACAGCCCTTACAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1827  
QY 180 AlaIleThrSerGlyLysGluValLeuGlyArgHlStrSerValIleGluValAsp 199  
DB 1828 GCTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1887  
QY 200 GluValArgArgLeuValTyrllePheGluGlyThrLysAspSerProLeuGlnHlStr 219  
DB 1888 GAGAGACCAAGCTGTGTGATCTTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1947  
QY 220 TyrlleValSerTyrlleAsnProGlyGluValThrArgLeuThrAspArgGlyTyrlle 239  
DB 1948 TAGCTGTACAGTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2007  
QY 240 HlSerCysGlyLeuSerGlnHlStrCysAspPhePheIleSerLysTyrlleSerValGlu 259  
DB 2008 CATAGCTCTTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2067  
QY 260 AsnProHlStrValSerLeuTyrlleLysSerProGluAspArgProThrCysLys 279  
DB 2068 ACGCCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2127  
QY 280 ThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrllePro 299  
DB 2128 CAGCCCGCTTCTGGGCTGACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2187  
QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrlleMetLeuTyrlle 319  
DB 2188 CCGAGATCTTCATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2247  
QY 320 ProHlAspLeuGluProGlyLysLysTyrlleProThrValLeuPheIleTyrlleGlyPro 339  
DB 2248 CCCACGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2307  
QY 340 GlnValAlaIleAlaGlyAlaProValThrLeuTyrllePhe 353  
DB 2308 CAGTGTGACAGTGTGATTAATCTTCAAGAGATCAAGTACTTGGGCTCAACACAGAG 2367  
QY 354 TyrllePheThrGlyTyrlleThrValThrValThrValThrValThrValThrValThr 359  
DB 2368 GCTCCCTGAGTACCCGCTGTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2427  
QY 360 GluArgTyrlleMetGlyHlStrProAspGlnAsnGluGln 371  
DB 2428 CGGTTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2487  
QY 371 371  
DB 2488 GGCCTGACATTCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2547  
QY 372 GluTyrlleGlySerValAlaMet-GlnAlaGluLysPheProSe 387  
DB 2548 GGTGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2607  
QY 387 rgluProAsnArgLeuLeuLeuLeuHlStrGlyPheLeuAspGluValHlStrPheAlaHl 407  
DB 2608 AAGGCGCAACCGCTTATCTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2667

```

OY 407 sThrSerIleuLeuSerPheLeuValArgAlaGlyLysProTyrAspLeuGlnIleTy 427
Db 2668 CACAACCTTCTCGTCTCCCACTGATCCAGCAGGAAACCTTACCACTCCAGATCTA 2727
OY 427 rProGlnGluArgHisSerIleArgValProGlnSerGlyGlnHisTyrGlnLeuHisLe 447
Db 2728 CCCCACGACGAGACACAGTATTCGCTGCCCGAGTCGGCGAGCACATATGAAGTCACGCTT 2787
OY 447 uLeuHisTyrLeuGlnGlnLeu 455
Db 2788 GCTGCACCTTCTACAGGAATACCTC 2812

```

Search completed: October 16, 2003, 09:09:25  
 Job time : 397.655 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 15, 2003, 17:24:12 (Search time 2793.69 Seconds)

(without alignments)  
4045.398 Million cell updates/sec

Title: US-10-070-464-5  
Perfect score: 2482  
Sequence: 1 TGTANPKVTRKMSSEIMDAE.....HLHYIGENGSRIALAKVI 465

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh  
-O=/cgr2\_1/USPTO.spool/US10070464/runac\_15102003\_113553\_24823/app\_query.fasta\_1.2652  
-DB=EST -QFMT=fastap -SUFFIX=xet -MINMATCH=0.1 -LOOPCH=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsm62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10070464@CCN 1.1.8056@runac\_15102003\_113553\_24823 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

#### Database:

EST: \*  
1: em\_estba: \*  
2: em\_estbm: \*  
3: em\_estin: \*  
4: em\_estml: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: em\_gse\_hum: \*  
18: em\_gse\_inv: \*  
19: em\_gse\_pln: \*  
20: em\_gse\_vit: \*  
21: em\_gse\_fun: \*  
22: em\_gse\_mam: \*  
23: em\_gse\_mus: \*  
24: em\_gse\_pro: \*  
25: em\_gse\_rod: \*  
26: em\_gse\_png: \*  
27: em\_gse\_vrl: \*  
28: gb\_gse1: \*

29: gb\_gse2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2326	93.7	3143	11 AK016546	AK016546 Mus muscu
2	2234	90.0	5517	11 AK029788	AK029788 Mus muscu
3	1488.5	60.0	3327	11 AK050021	AK050021 Mus muscu
4	1488.5	60.0	3376	11 AK050023	AK050023 Mus muscu
5	1478.5	59.6	3457	11 AK078301	AK078301 Mus muscu
6	1397	56.3	910	13 BX372276	BX372276 Mus muscu
7	1265.5	51.0	853	12 B1223892	B1223892 602941035
8	1202	48.4	993	10 BG259714	BG259714 602380072
9	1157.5	46.6	1021	12 BG974587	BG974587 602844820
10	1127.5	45.4	774	14 CB233750	CB233750 AGENCOURT
11	1117.5	45.0	1042	12 BM557438	BM557438 AGENCOURT
12	1082	43.6	1041	12 B1084090	B1084090 602869453
13	1063	42.8	762	14 CD352521	CD352521 UT-M-G10
14	1055	42.5	753	9 AL040398	AL040398 DKF2P334A
15	1040	41.9	831	12 BG974324	BG974324 602844075
16	1034.5	41.7	957	13 BQ675006	BQ675006 AGENCOURT
17	1028	41.4	956	14 CB153590	CB153590 K-EST0211
18	1015	40.9	910	13 BQ675260	BQ675260 AGENCOURT
19	1015	40.9	968	13 BQ671635	BQ671635 AGENCOURT
20	1011	40.7	897	13 BU916138	BU916138 AGENCOURT
21	981.5	39.5	1085	9 AL582206	AL582206
22	972	38.2	575	10 BG086878	BG086878 H3131F11
23	946	38.1	896	10 BF784154	BF784154 602108015
24	912	36.7	921	13 BU147109	BU147109 AGENCOURT
25	880.5	35.5	1082	13 BU239476	BU239476 60332338
26	864	34.8	682	10 BM473819	BM473819 BB473819
27	857	34.5	1090	12 BM10838	BM10838 AGENCOURT
28	850	34.2	853	10 BF694553	BF694553 602081767
29	849.5	33.2	845	13 BU216915	BU216915 603109373
30	831	33.5	469	9 AA465309	AA465309 aa24b08.x
31	811.5	32.3	943	13 BQ895618	BQ895618 AGENCOURT
32	802.5	32.3	888	13 BU145896	BU145896 AGENCOURT
33	797	32.1	658	10 BE388695	BE388695 601286345
34	792	31.9	467	14 CA541559	CA541559 C0607E01
35	760	30.6	789	10 BG709118	BG709118 602675382
36	746	30.1	785	12 B1148072	B1148072 602912419
37	719	29.0	674	10 BE888655	BE888655 601513061
38	712	28.7	746	14 BY751026	BY751026 BY751026
39	709.5	28.6	902	13 BQ642814	BQ642814 AGENCOURT
40	690.5	27.8	855	12 B1084885	B1084885 602869453
41	682.5	27.5	727	12 B1855677	B1855677 603383206
42	678	27.3	735	13 BU447277	BU447277 603766685
43	673.5	27.1	970	10 BG403555	BG403555 602419260
44	669.5	27.0	788	10 BG163397	BG163397 602338360
45	663	26.7	481	10 BF729761	BF729761 mab75g12.

#### ALIGNMENTS

RESULT 1  
LOCUS AK016546 3143 bp mRNA linear HTC 05-DEC-2002  
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:14932434P09 product:DIPEPTIDYL PEPTIDASE 8 homolog  
ACCESSION AK016546  
VERSION AK016546.1 GI:12855334  
KEYWORDS (Homo sapiens), full insert sequence.  
SOURCE HTC; CAP trapper.  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20499374  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komano, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel sequencer

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20530913  
11076661  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Komano, H., Aachi, T., Fukuda, S., Aikawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Plischmann, P., Gaasterland, T., Gissi, C., King, B., Kochiya, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stadhil, P., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Baren, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seye, T., Shibata, Y., Storch, K. F., Suzuki, H., Togo, Oka, K., Wang, K. H., Wetz, C., Whitaker, C., Williams, L., Wyszewski, B., A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohitsuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
21085660  
11217851  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3143)  
Adachi, J., Aizawa, K., Akahira, S., Akiyama, T., Arai, A., Aono, H., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanaizaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komano, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suuki, H., Tagami, M., Tagawa, A., Takehashi, F., Tanaka, T., Tejima, Y., Togo, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Direct Submission

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
Submitted (10-JUL-2000)  
Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

**COMMENT**

EXPLORATION Research Group RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp.  
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGAAGACATCCCAAGACCCTCTTTTCTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. Second  
strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAAGATTTCGAGTTAATTAAATTAAATCCCCCCCCC 3']. cDNA was cleaved  
with BamHI and XhoI. cDNA of size compressed longer than 7 kb was  
selected before cloning. Vector: a modified pluscript KS(+) after  
bulk excision from lambdaB FLC I. Cloning sites, 5' end: SalI, 3'  
end: BamBI. Host: DH10B.

**FEATURES**

Source	Location/Organisms
	1..3143
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="CS7BL/6J"
	/db_xref="FANTOM DB:4932434F09"
	/db_xref="MG1:1907554"
	/db_xref="taxon:10090"
	/clone="4932434F09"
	/sex="male"
	/tissue_type="testis"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="adult"
	327..3505
CDS	/note="unnamed protein product; DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens] (SPR_09HBMS, evidence: FASTY, 95.%ID, 100%length, match=2649) putative" /codon_start=1 /protein_id="BAB30295.2" /db_xref="GI:26382128" /db_xref="MG1:1921638" /translation="MKPIPSGRCNMAAMETQGLVEIFETAECEGNESQDRKLEPP FYVRYERSQLKKLADTRKHGYMMAKAHDPMFKTPDPSPHSRVYUAMSNGN RENTLFYSBILPXTINRAIVLMISMKPLDFOATLDGVNSREBELREKRIGTVGI AAYDHGSGSTPLFOAGSGIVYLIKDGPHGTQOLRPVNVPSSCPIMDKPLCPAD PDNIARTHSNDWITSNLYTREERRITYNHMLNMEDPRASAGATTVLOBEPHYSG YMWCQAKERTPSGGKILRLIYEENDESVEVIHTVSMTLRADSRFRYPKTGANK VTREKSEIIVDDAAGIIVDIKDENVQFEILFGCEVIYIAAGTWPEBKHAMSLINDKS OTHLQVLISPELPFVEDMDARORLIESVPDSVPLLIYEETDIDWINIHDFHF POHQHEIDIEFIIPASECTGFRHLKYKITSLKSEKYKSSGGLPAPSDPCPKIEELTI TSGEEMTLHGSGNISWDERKLIVFEGTDSPLEHLIYTSVANPEVVRLDRGVS HSCCLSHDCDFPISKYSNOKNPCHCYLYKTSDEDDPVHTKTFPMATILLDSAPLIPY TPREIRSFESTTGFTLYXMKLYKPHDLOPKKRTTVLFYGPOVOLYNMFQKTKFR LNLTLASIGVIVYIDNRKSGCHKLKEFGARKYVMGOILEIDDQVEGLQVLA5YDPDL DRVGHWSYGVGLISMALMQRSIDIFRWLAGAPVAFHLSIVDTGYTERVGHGPDNEQ GYVIGSYVAQAERKPSEPSPNRLLLHGFGLDENYFAHTSIILSELVAGKPYDLOIYPO ERSHRIPESGEHYEHLHLHYLEOENGSRIALVKVI"

**BASE COUNT**

694 a	688 c	773 g	788 t
-------	-------	-------	-------

**ORIGIN**

Alignment Scores:	Pred. No.:	Length:	3143
Score:	2326.00	Matches:	443
Percent Similarity:	80.18%	Conservative:	10
Best Local Similarity:	78.41%	Mismatches:	12
Query Match:	93.71%	Indels:	100
DB:	11	Gaps:	1

OS-10-070-464-5 (1-465) x AK016546 (1-3143)

Db	1308	ACGGCACAAGCAAAACCCAAAGCTCACTTTCAAGATGTCGGAGATTGTTGTAAGCTGCA	1367
Oy	21	GIYAAGLIEIIEASPVALLIeASPLySgIueuIeGINProPheGIuIleuPheGIu	40
Db	1368	GGAGGGATTATAGATGCTACATAGATTAAGAACTGGTTCAACCTTTGCAATTCGTGTTAG	1427
Oy	41	GIYVAGIuTYrIIaIaIaYgaIaGIyTPThrProGIuGIyVeyTYrAlaITPserIle	60
Db	1428	GGAGTGTGAATATATATGGCCAGACCCGATGAGACTTCAGAGGGAAAAACAGCTGGTCCATTC	1487
Oy	61	LeuLeuASpASgSerGINthzrYgIueuGINIleValIeSerProGIuIleuPheIle	80
Db	1488	CTACTAGACCGTTCACAGACTCACTCGCAGATAGTTGATGATCCCTCGAETATTCATC	1547
Oy	81	ProValGIuASpASpValweGIuYrGIuYrIaIeGIuSerValProASpSerVal	100
Db	1548	CCAGTAGAAGAGATGATGCATGACACAGACAGACTTAATAGATCGAGTTCCTGACTCTGTG	1607
Oy	101	ThrProIeUIeIIErTYGIuGIuYrThrASpIIErPIIeASpIIeHIAspIIePhe	120
Db	1608	ACACCACTGATCATATATGAAMAAACAACAGCACTGTGATTAATATCCAGATATATTTT	1667
Oy	121	HisValPheProGINserHIEGIuGIuGIuIeGIuPheIlePheAlISerGIuCyIeYs	140
Db	1668	CATGTTTTTCCCAACTCATGAATGAATGAATGATTTATTTTGGCTCGATGATGCAA	1727
Oy	141	ThrGIyPheATgHIEuTYrIySIErSIIeIeulYegIuSerIeTYrIyAATg	160
Db	1728	ACAGGTTTTCCGTCATCTGATATAAATCAATCCATTTTAAAGAGACAAATATAAACGG	1787
Oy	161	SerSerGIyGIyIeProAlaProSerASpPheIySASpIIEIyGIuGIuIleIa	180
Db	1788	TTCAGTGGIIGACTACTGCCCCCAAGTGTTCGAATGCTCATCAAGAAAGAAATTAACA	1847
Oy	181	IIErThrSerGIyGIuTYrPIuIValIeGIyAtrgHISgIySerASpIIEGINValASpGIu	200
Db	1848	ATTACCACTGGTGATGGGAAGTACTTGCCCGCATGATCTAATATCTGGGTTGATGAA	1907
Oy	201	ValATgATgIeUValTYrPheGIuGIyThrIyASpSerProIeGIuHIEIleTYr	220
Db	1908	GCCAGAAAGCTGGTGTACTTTGAAAGCACAAAGACTCTCTTTGGACATCACCTGTAC	1967
Oy	221	ValValSerTYrValASpProGIyGIuValThrATgIeUThrASpARGIYTYrSerHIS	240
Db	1968	GTGACCACTATGCAAAACCTGGAGAACTGTGAGAGCTGACCTGACCGTGGCTACTCAAC	2027
Oy	241	SerCYsCYSIeSerGINHISCyASpPhePheIleSerIySrySerASpGINIyASpN	260
Db	2028	TCTGCTGCTGCACGCCGACATTGTGACTTCTTCATTAAGTAAAGTACAGCAACAGAAAT	2087
Oy	261	ProHISCySValSerIeUTyIyIeUSeSerProGIuASpASpProThCyIySthr	280
Db	2088	CCACACTGTGTGCTCCCTCTACAAACTCCAAAGTCTGTGAGATGACCCCGTTCATTAACA	2147
Oy	281	LYSGIuPheTPAlaThrIIEuIeUASpSerIaGIYProIeUProASpTYrThProPro	300
Db	2148	AAGGAATTTTGGCCACCACTTTGGATTCAAGAGGTCTCTTCTCGACTACAAACCCCTTCA	2207
Oy	301	GIuIlePheSerPheGIuSerThrThrGIyPheThrIeUyrgIyMetIeUyTYrIyASp	320
Db	2208	GAATTTTTTCTTTGAAGAATACACTGATTTACACTGATNGAATGTTGTATTAAGCTT	2267
Oy	321	HisASpIeUGINProGIyIyIeUyIeTYrProThrValIeUbeHIErTYrGIyGIyProGIN	340
Db	2268	CATGACCTPACAACCTGGAAAGAAATACCCCACTGTATTATTCATATATGAGTGCTCCAG	2327
Oy	340	-----	340
Db	2328	GTGACACGTGTGAACAATCGGTTTAAAGAGATCAAGTATTTCCGCCCTGAAACACCTGACC	2387
Oy	340	-----	340

[illegible]



Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashtwagi, K., Fujitake, S., Inoue, K., Togawa, M., Ozawa, M., Ohata, E., Watehiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system—384-Format sequencing pipeline with 384 multichannel sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

QY 262 HiscysValSerLeuTyrIleuSerSerProGluAphAapProThrCysIysThrIys 281  
 Db 3923 CACTGTGTGTCTCCCTCAAACTCTCAAGTCTAGATGACCAAGTTTCAATAAACAAG 3982  
 QY 282 GluPheTpaIaThrIleLeuApsSerIaGlyProLeuProAapTyrThrProGlu 301  
 Db 3983 GAATTTGGGCGACCATTTTGATTCAGAGGTCTCTTCTGACTACACCCCTCCAGAA 4042  
 QY 302 IlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrIysProHis 321  
 Db 4043 ATTTTCTTTTGAAGTACTACTGAGATTACACTGATGAGAAATGTGTATAGCCCTCAT 4102  
 QY 322 AspLeuGlnProGlyIleuIysTyrProThrValLeuPheIleTyrGlyIysProGln 340  
 Db 4103 GACCTACACCTGGAAAGAAATACCCACGCTGTTATTCATATATGAGTGTCCACAGTC 4162  
 QY 340 ----- 340  
 Db 4163 CAGCTGTGAACAATCGTTTAAAGAGTCAAGTATTTCCGCCCTGAACCCCTGCCCTCC 4222  
 QY 340 ----- 340  
 Db 4223 CTGGGTTATGTGTGTGTGATAGACACAGGGGATCTGTACCGAGACTTAAATTT 4282  
 QY 340 ----- 340  
 Db 4283 GAAGCGCCCTTTAAATATTAATAATGGGTCAAAATAGAAATCGATGATCAATGGAAGACTC 4342  
 QY 340 ----- 340  
 Db 4343 CAGTACTAGCATCTGATATGACTTGAATGATGAGTGGGATCCACCGCTGG 4402  
 QY 341 -----Val 341  
 Db 4403 TCCTATGNGGCTACCTCTCCCTGATGACATTAATGACAGAGTGGATATCTTCCGGGTG 4462  
 QY 342 AlaIleAlaIyAlaProValThrLeuTyrIlePheTyrAapThrGlyTyrThrGluAarg 361  
 Db 4463 GCTATTCTGGGGCCCGGACGACACCTCTGATCTCTTATATACAGATACACGAGAGCGC 4522  
 QY 362 TyrMetGlyIleAProAapGlnAargGlnGlyTyrTyrLeuGlySerValAlaMetGln 381  
 Db 4523 TATATGGGTACCCCTGACCAAGATGAAACAGGGCTACTAGATGAGTCTGTGGCATGCA 4582  
 QY 382 AlaGluIysPheProSerGluProAapAargLeuLeuLeuLeuIleGlyPheLeuAapGlu 401  
 Db 4583 GCGGAGAGGTTCCTCCAGAACCAACCGGTACTCTCTTGCATGGGTCTTGGATGAG 4642  
 QY 402 AsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAargAlaGlyIysPro 421  
 Db 4643 AATGTTCACTTTCACACACACCATATATTCGTAGTTTTTTAGTGAAGGCTGGAAAGCCA 4702  
 QY 422 TyrAapLeuGlnIleTyrProGlnGluAargHisSerIleAargValProIleSerGlyGlu 441  
 Db 4703 TATGACTTACAGATCTATCTCTCAGAGAGGACACAGCATCAGAGTTCTGTGAGGA 4762  
 QY 442 HisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAapLeuGlySerAargIleAla 461  
 Db 4763 CACTATGAACTGACACCTGCTCCACTACTCTCAGAGAACCTTGATGCGTATGCGTCT 4822  
 QY 462 LeuIysValIle 465  
 Db 4823 CTCGAAAGTAT 4834

RESULT 3  
 AK050021  
 LOCUS  
 DEFINITION  
 AK050021  
 VERSION  
 AK050021.1  
 KEYWORDS

3327 bp mRNA linear HTC 05-DEC-2002  
 Mus musculus adult male liver tumor cDNA, RIKEN full-length  
 enriched library, clone: C7300003D12 product: D1PEPRTD1L PEPRTDASE 9  
 homolog (Homo sapiens), full insert sequence.  
 HTG; CAP trapper.

SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 2  
 99279253  
 10349636  
 3  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 4  
 20499374  
 11042159  
 5  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 6  
 20530913  
 11076861  
 7  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,  
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsumoto, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
 Quackenbush, J., Schiraldi, L.M., Staubli, F., Suzuki, R., Tomita, M.,  
 Wagner, L., Washio, T., Sakai, K., Okada, T., Furuno, M., Aono, H.,  
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
 Fletcher, C., Fujita, M., Gariboldi, M., Gietz, S., Hill, D.,  
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,  
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
 Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,  
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S.  
 and Hayashizaki, Y.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409 (6821), 685-690 (2001)  
 8  
 21085660  
 11217851  
 9  
 The PANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)



QY	340	-----	340
Db	2234	CACCTTCAGAGGGGCGCCCTGAAAAATCAAAATGGCCAGGTGAGACATTGAGACACCGAGTGGAA	2293
QY	340	-----	340
Db	2294	GCGTTGCAGTACGTGGCTGAGAAAGATGAGCTTCATTGACTTGAGCCGAGTGGCATTCAT	2355
QY	340	-----	340
Db	2354	GCGTGGTCCATACGGCGGCTTCCTCACTCATATGGGCTCATTCACACAGCCCAAGTGTTC	2413
QY	341	---ValAlaIleAlaGlyAlaProValThrLeuPrlIlePheTyrAspThrGlyTyrThr	359
Db	2414	AAGGTAGCCATTGGCGGCGCTCTCTCACTGTGTGATGGCCTTATGACACAGGGTACAG	2473
QY	350	GIuArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrTyrLeuGlySerValAla	379
Db	2474	GAAAGATACATGATGATGTCCCGGAAATTAACACAGCAAGGCTATGAGGCGAGGTGTGTACC	2533
QY	380	MetGlnAlaGlyIuValPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeu	399
Db	2534	CTGCATGTGGAGAACTCCGCCATAGGCTTAACGCCCTTATCTCCACAGGCTTCCTG	2593
QY	400	AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly	419
Db	2554	GACGAGAAAGCTTCACCTTCTTCCACACAAATTTCTCGTGTGCCAGCTATTCGAGACAGGA	2653
QY	420	LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer	439
Db	2654	AAGCCATACACAGCTTCAGATCTACCCCAACGAGAGACATACATCCGCTGCGGAGTCC	2713
QY	440	GlyGluHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlnGluAsnLeu	455
Db	2714	GAGAGCATTTACGAGGTGACGCTGCTGCACTTCTGCAGAAACACTTG	2761
RESULT 4			
AK050023			
LOCUS			
DEFINITION			
AK050023			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Oikawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabili, F., Suzuki, R., Tomita, M., Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Bersh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bul, C., Fletcher, C., Fujita, M., Gaibolli, M., Gustincich, S., Hill, D., Hochmann, L., Hume, D.A., Kamaya, M., Lee, N.H., Lyons, P., Marchionni, L., Meshima, J., Mazzarelli, J., Mombere, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Togo, Oka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L., Wyszynski, B., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlschütter, S., and Hayashizaki, Y.

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3376)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imochi, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Okach, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsumura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sato, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashi, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

1  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers

FEATURES  
SOURCE

1. 3376  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:C730003H19"

```

/db xref="taxon:10090"
/clone="C730003H19"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
235..2824
/note="DIPEPTIDYL PEPTIDASE 9 homolog (Homo sapiens)
(SPTRIAL47179, evidence: FASTA, 92%ID, 99.6%length,
match=2580)
misc_feature

```

BASE COUNT 698 a 1046 c 932 g 700 t  
ORIGIN

## Alignment Scores:

```

Pred. No.: 7.4e-153 Length: 3376
Score: 1488.50 Matches: 278
Percent Similarity: 62.59% Conservative: 70
Best Local Similarity: 50.00% Mismatches: 107
Query Match: 59.97% Indels: 101
DB: 11 Gaps: 2

```

US-10-070-464-5 (1-465) x AK050023 (1-3376)

```

QY 1 ThrGlyThrAlaAspProLysValThrPheLysMetSerGluLeuMetIleAspAlaGlu 20
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1154 ACAGGCGACGAAGAACCCCAAGATGCGCTGAGCTGAGCTCCAGAGGACATCAG 1213

QY 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu 40
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1214 GGCMAAATCGTGCTGACCTGAGAGAGAACCTGTGATACGCACTTCCCTTTTCCC 1273

QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1274 AAAGTGAAGTACATCGCCCGGCTGCTGACACAGGACCGCAAAATGCTGGGCGCATG 1333

QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1334 TTCCTGAGCCGTCCTCCAGCAACGGCTTCAGCTTGTCTCTGCGCCCTGCTCTTCATC 1393

QY 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1394 CCGGCGCTTGAAGTACAGGCGGACGACGACCTCCAGAGCGCTGCCCAAGAAATGTG 1453

QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1454 CAGCCTTTGTCTATGAAAGATGACCAATGTCTGATCAACGTCCACGACATCTTC 1513

QY 121 HisValPheProGlnSerHis--GluGluGluIleGluPheIlePheIleAspGluCys 139
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1514 CACCGGTTTCTCAGGCTGAGGCGCCACAGACATTTGTTTCTTCGTGCCAAGAAATGC 1573

QY 140 LysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1574 AAACATGCTTCTGCGCACCTGTACAGGATCAAGTGAACCTTAAACCAAGAGCATAGAC 1633

QY 160 ArgSerSerGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIle 179
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1634 TGAGCGAAGACCCCTACCGCTTACAGAGATGAGTTTAAGTCCCAATCAAGAGAGAGCTC 1693

QY 180 AlaIleThrSerGlyLysTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1694 GCGCTGACGATGCGGATGAGGATGCTTGTGCGGATGATGCTCAAGATCTGGGCTCAAC 1753

QY 200 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisLeu 219
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1754 GAGGACGACGAAGCTGCTACTTTCAAGATACAAAGACACCGCTGGAACATCAGCTC 1813

QY 220 TyrValValSerTyrValAspProGlyGluValThrArgLeuThrAspArgGlyTyrSer 239
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1814 TATGTGTGCTACGATCAGATCAGAGCGGAGATGCTGCGGCTACACACGCTCGGCTTCTC 1873

```

```

QY 240 HisSerCysAlaSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1874 CACAGCTGCTCCATGAGCCAGAGCTTCGACATGTCGTGAGTCACTACACAGCTGAGC 1933

QY 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspPheProThrCysLys 279
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1934 ACGCCACCTGTGTATACATGTGTCAAGCTGAGCGGCCCATATATGACCCATGCAAG 1993

QY 280 ThrLysGluPheThrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 299
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 1994 CAACACGCTTCTGGGCGACATGATGAGGACGCCAATGCCCCCAGCATATGTCGCC 2053

QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 319
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2054 CTGAGATCTTCCATCTCCACATCCCGTGCAGAGCTGACGCTTACGCGCATGATCTCAAG 2113

QY 320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 339
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2114 CCACACACCTGTGACCTGAGGAGGAGACCCCATCTGCTTGTGTATGGGCGCCA 2173

QY 340 Gln----- 340
DB 2174 CAGGTGATGTGTGTAACAACCTCTTAAGGATCAAAATCCTGCGGCTAAATACACTG 2233

QY 340 ----- 340
DB 2234 GCATCTTGGGCTATGCTGTGTGTGATGATGATGTCGGGCTCTGTTCAGCGGCGCTG 2293

QY 340 ----- 340
DB 2294 CACTTGAGGGGGCCCTGAATAATCAATGGCCAGGTGAGATGAGACAGGTGAA 2353

QY 340 ----- 340
DB 2354 GCGTTCAGTACGTGCTGAGAGATATGCTTATTGACTTGAAGCCGATGCCATTCAT 2413

QY 340 ----- 340
DB 2414 GCGTGTCTACAGCGGCTTCTCTGATCATGAGGCTCATCCACAAGCCACAAGTGTTC 2473

QY 341 ---ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 359
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2474 AAGTATGACCTTCTGAGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 2533

QY 360 GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2534 GAACGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2593

QY 380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeu 399
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2594 CTGCATGTGAGAGAAAGCTGCCAATGAGCTTAAACCGCTGCTTATCTCCACAGGCTTC 2653

QY 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGly 419
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2654 GACGAAACCTTACTTCTTCCACACAAATTTCTGTGTGTCAGCTGATCCAGAGACGA 2713

QY 420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2714 AAGCATACACAGCTTCAAGATCTACCCAAAGAGAGACATAGATCCGCTGCCGAGTCC 2773

QY 440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeu 455
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 2774 GGAGAGCATTAACGAGGTGAGCGTGTGACCTTCTTCAGAGAACACCTG 2821

RESULT 5
AK078301 3457 bp mRNA linear HTC 05-DEC-2002
LOCUS DEFINITION
MUS musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430584G1 product:DIPEPTIDYL PEPTIDASE 9
homolog (Homo sapiens), full insert sequence.
ACCESSION AK078301
VERSION AK078301.1 GI:26347124
KEYWORDS HTC; CAP trapper.

```

SOURCE ORGANISM	Mus musculus (house mouse)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meeth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
REFERENCE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
REFERENCE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, Y., Nishi, K., Kitsuma, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
REFERENCE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kaubawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Knehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotamoni, L., Hume, D. A., Kamuya, M., Lee, N. H., Lyons, P., Mochizuki, L., Mashima, J., Mazzarelli, J., Mombaeers, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
REFERENCE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 3457)
PUBMED	Adachi, J., Aizawa, K., Akiyama, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashimoto, M., Hayashizaki, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kaich, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

**TITLE**  
**JOURNAL**

**COMMENT**  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration and Chemical Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-ree@cc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in RIKEN.  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/  
 Location/Qualifiers

**FEATURES**  
 source

1. 3457  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:6430584G11"  
 /db\_xref="taxon:10090"  
 /clone="6430584G11"  
 /sex="male"  
 /tissue\_type="olfactory brain"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 182..2770  
 /note="unnamed protein product; DIPEPTIDYL PEPTIDASE 9  
 homolog [homo sapiens] (SPR|AAL47179, evidence: FASTV,  
 92%ID, 99.6%length, match=2580)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC37211.1"  
 /db\_xref="GI:26347125"  
 /translation="MCSGVSPVEQVAAGMDMTAARPCVQKSHMDGLSHIIGSRKSS  
 GLTYSKAPHDFOFQKDESGPHSHRLYYLGMPPGSENSLTYEIPKVKKALLLL  
 SMKMDLHFOATPHHGYVSRBEELRRKRLKGLVSDYHSSGLFIPQANSLLFPH  
 CROGKMGPMVSPMKPLEIKTQCGSPMRDKICADPAFSPFINNSDLVAVNITGEE  
 RLRTFCQSGAGVLDNFKSGAVATFVIOEDRDITGCMCPTASBSBGKTLRL  
 EEDVESSEVYIHQSPALERRKTDSPYPRGSKNPALKLAEALQDHQKIVSCEE  
 KELVQPSLSFPKVEYLARAGKTRDGYAAMFLYRQQLVLLPLAFLPAVESEE  
 AGRARARAVKQVQVPIVEEYLVNINVDIHPHPOAGQODFCFLANESKGL  
 KLYRIVVELKTDYDWEPLSLPTDEDFKCPDKKEVALTSGEMVLRSHGSKIVNED  
 TKLYVRCGYKDTPLREHLLVYVSYESAETIRLTLTGSHSCSMQSDPMPSHYSSVS  
 TPFCVHYKLSGPDDPLHAKQPRFWMASMEANCPDYVPELIFHFTRADVLYGMN  
 YKPTLQPGRRKHPFLVLFVYGQPOVLVNNSEFKGIKYRLNTLAEALGYAVVYIDRGSG  
 QRGHGFKAALKNQMGQVEIEDQVGLYVAEKGYFIDLSRYAINGMSYGGFLSMGLI  
 HKPQRFVAILAGAVTVMAVYDGTGTYRVMVDVPPNNKQGYEAGSVALLHEKLPNEP  
 LLILHGLDENVHFHFTNPLVSQLIRAKRKYQLQIYRNEHNSIRCRSEGEHYETLLH  
 FLEGHL"

**BASE COUNT** 766 a 1033 c 926 g 732 t

**ORIGIN**

**Alignment Scores:**

Prd. No.:	9.82e-152	Length:	3457
Score:	1478.50	Matches:	277
Percent Similarity:	62.41%	Conservative:	70
Best Local Similarity:	49.82%	Mismatches:	108
Query Match:	59.57%	Indels:	101
DB:	11	Gaps:	2

US-10-070-464-5 (1-465) x AK078301 (1-3457)

Oy 1 ThrcglYThralaansnProlysvaLThPhelysmetSerglullemetllaaspalaglu 20  
 Db 1100 ACAGGCGACGAAGAACCCCAAGATTGCCCTGAAGTGTCTAGCTCCAGAGACCAACATCAG 1159  
 Oy 21 GLYArgllelleaspaVallleasplygcluleuileglnProphcgululeuPhcglu 40  
 Db 1160 GGCMAATTCGTTCAAGCTCGCAAGAGAACTGGTACAGCAATTCCTCTTTTCC 1219



QY 41 GlyValGluTyrIleAlaArgAlaGlyTyrPheProGluGlyLysTyrAlaTyrSerIle 60  
 DB 1220 AAGTGTGAGTACATCCCGGGCTGGCTGGACACGGGACGCAAAATATGCTGGGCGCATG 1279  
 QY 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValIleuIleSerProGluLeuPheIle 80  
 DB 1280 TTCCTTACCGTCCCGACCAACGGCTTCGCTCTCCACCCCTGCTCTCTTCATC 1339  
 QY 81 ProValGluAspArgValIleArgGlnArgLeuIleGluSerValProAspSerVal 100  
 DB 1340 CCGGCGGTGAGTGAAGGCGCCAGCGGACGGCAGCTGCCAGAGCCGTCGCCAAGAAATG 1399  
 QY 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 DB 1400 CAGCCCTTTGTCTATATGAGAAGTCAACCAATGCTGATCAACCTCCAGCATCTTC 1459  
 QY 121 HisValPheProGlnSerHis---GluGluGluIleGluPheIlePheAlaSerGluCys 139  
 DB 1460 CACCCGTTCTCAGGCTGAGGCGCAGGACCTTTGTTTCTTCGTCGCAAGATGC 1519  
 QY 140 LysThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLys 159  
 DB 1520 AAGACTGGCTTCCGACCTGTACAGAGGTACAGTGAACTTAAACCAAGACTATGAC 1579  
 QY 160 ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIle 179  
 DB 1580 TGGACCGAACCCTCAGCCCTACAGAAAGATTTAAGTCCCATCAAGAGAGAGTGC 1639  
 QY 180 AlaIleThrSerGlyGluTyrGluValLeuGlyArgHisGlySerAsnIleGlnValAsp 199  
 DB 1640 GGCCTACACAGTGGCAAGTGGAGGCTCTGTCCAGAGGATGGCTCCAAAGCTCTGGCTCAAC 1699  
 QY 200 GluValArgArgLeuValTyrPheGluGluTyrThrLysAspSerProGluGlnHisIleu 219  
 DB 1700 GAGCACAAGAGTGGTGTATCTTCAAGGTAAGAAAGACACCGCTGGAACTACCTC 1759  
 QY 220 TyrValIleSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 239  
 DB 1760 TATGTGTAGTACAGACGACGACAGGAGATGCTGGGCTCCACACGCTCGGCTCTC 1819  
 QY 240 HisSerCysAlaSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 259  
 DB 1820 CACAGTGTCTCCATGAGCCAGACCTTCGACATGTTCTGTGCTACTACACGACTGTGAGC 1879  
 QY 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 279  
 DB 1880 AGCCACCCCTGTGTATGTATGTAAGCTGAGCGCCCGCATATGACCCACGACGACAG 1939  
 QY 280 ThrLysGluPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 299  
 DB 1940 CAACACAGCTTCTGGGCGACAGCATATGAGAGCAGCCAAATGCCCCAGACTATGAGCC 1999  
 QY 300 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 319  
 DB 2000 CCGAATATCTTCCATCTCCACACCCGCGACAGCTGAGCTACAGCGCAAGATCTCAAG 2059  
 QY 320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLysPro 339  
 DB 2060 CCAACAACCTGCAACTGGAGAGAAAGACCACTGTGCTCTTTGTCTATGAGGAGGCCA 2119  
 QY 340 Gln----- 340  
 DB 2120 CAGGTGACAGTGTGAACAACCTCTTAAAGGCGCATCAATACCTCGGCGTAAATACAGT 2179  
 QY 340 ----- 340  
 DB 2180 GCATCTTGAGGCTAATGCTGTGTGTGATGATGATGTCGGGCTCTGTCTACGCGGAGCTG 2239  
 QY 340 ----- 340  
 DB 2240 CACTTGAGAGGGGCGCTGAATAAATCAAAATGAGGCGAGTGTGAGATTGAGGACCAAGTGA 2299

QY 340 ----- 340  
 DB 2300 GGCTTGACATAGTGAGCTGAGAAATATGAGCTTATTGATGAGCGGAGTGCATCAT 2359  
 QY 340 ----- 340  
 DB 2360 GGCTGTCTTACGGCGCTTCTCTTCACTCATGAGGCTCATCCAGACCAAGTGTTC 2419  
 QY 341 ---ValAlaIleAlaGlyAlaProValThrLeuTyrIlePheTyrAspThrGlyTyrThr 359  
 DB 2420 AAGGTAGCATTCGCGGCGCTCTCTGACTGTGTGATGAGTGCCTATACACAGGTTACAG 2479  
 QY 360 GluArgTyrMetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAla 379  
 DB 2480 GAACGATACATGATGATCTCCCGCAATAATACAGCAAGGCTATGAGCAGAGGTGTGAGCC 2539  
 QY 380 MetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeu 399  
 DB 2540 CTGCATGTGAGAAAGTGCACATGAGCCTTAACCGCTCTTATCTCCACGCGCTTCTG 2599  
 QY 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGly 419  
 DB 2600 GAGGAAAGCTTCACTTCTTCCACACAAATTTCTGTGTGTCAGCTGATCCAGACAGCA 2659  
 QY 420 LysProTyrAspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSer 439  
 DB 2660 AAGCCATACAGCTTCAGATCTACATCCCAAGAGAGACATAGCATCCGCTGCGCGAGTCC 2719  
 QY 440 GlyGluHisTyrGluLeuHisLeuLeuHisTyrTyrLeuGlnGluAsnLeu 455  
 DB 2720 GGAGACATATTACAGAGTGAAGCTGTGCTGCTTCTGACAGAACACCTG 2767  
 RESULT 6  
 BX372276/c  
 LOCUS BX372276 910 bp mRNA linear EST 08-MAY-2003  
 DEFINITION Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 ACCESSION BX372276  
 VERSION BX372276.1 GI:30448117  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Pukaryofa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1. W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7542.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BA1043ZH06.CS04080.1c1cluster=7542.r.  
 Contact: Feng Liang Email: fliang@life.techn.com URL:  
 http://fulllength.invitrogen.com/invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0BA1043ZH06\_CS04080\_1.  
 Location/Qualifiers  
 1. 910  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL005YD02"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V



BASE COUNT sites of the PCMVSPORT 6 vector. Library was normalized.  
 251 a 193 c 176 g 288 t 2 others

Alignment Scores:  
 Pred. No.: 1,24e-143 Length: 910  
 Score: 1397.00 Matches: 282  
 Percent Similarity: 96.91% Conservative: 0  
 Best Local Similarity: 96.91% Mismatches: 9  
 Query Match: 56.29% Indels: 4  
 DB: 13 Gaps: 0

US-10-070-464-5 (1-465) x BX372276 (1-910)

1 ThrGlyThrAlaAsnProlyValThrPheLeuMetSerGluLeuMetIleAspAlaGlu 20  
 869 ACAGGACAGCAAACTCTTAAGCAGCTTTTAAAGATGCAAAATAATGATGATGCTGAA 810  
 21 GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleAspGlu 40  
 809 GAAGAGATCATATGATCATAGTAAGAACTAATTCACCTTTGAGATTCATT-GAA 752  
 41 GlyValGluTyrIleAlaArgAlaGlyTyrThrProGluGlyLysTyrAlaTyrSerIle 60  
 751 GNAAGTGATATATATGCGAGAGCTGATGAGTCTCTGAGGGAATAATATGCTGTCATC 692  
 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle 80  
 691 CTACTAGATGCTCCCAAGACTCCCTACAGATAGTGTGATCCACCTGATTAATTTATTC 632  
 81 ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal 100  
 631 NCAGTGAAGATGATGTTATGAAAGCAGAGACTCATTTGACATGCTGATTCGTG 572  
 101 ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe 120  
 571 AGCCCACTATATCTATGAGAAACAACAGACATCGATAATATCCATGACATCTTT 512  
 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheIleAspGluCysGly 140  
 511 CATGTTTTTCCCAAGATCAAGAAAGAAATGATTTATTTTCCCTGATCAACAA 452  
 141 ThrGlyPheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArg 160  
 451 AAGAGTTCCGTCATTTATTAACAATTAATCTATTTTAAGGAAAGCAATTAACA 392  
 161 SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAla 180  
 391 TCCAGTGAGGCTGCTGCTCCCAAGTGAATTCAGATGCTCTATCAAGAGAGATAGCA 332  
 181 IleThrSerGlyGluTyrGluValIleGluYargHisGlySerAsnIleGlnValAspGlu 200  
 331 ATTACAGAGTGGAATGGAGAGTTCTTGCCGGCAGATGATTAATATCCAAAGTTGATGAA 272  
 201 ValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlnHisIleLeuTyr 220  
 271 GTCAGAAAGCTGTATATTTGAAGCACAAGAACTCCCTTTAAGCATCACTGTC 212  
 221 ValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHis 240  
 211 GATAGTACGTTACGTAATCTCGAGAGGTGACAAAGCTGACTGACCTGCTACTACAT 152  
 241 SerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGluLysAsn 260  
 151 TCTTGCTGATCAGTACGACTGTGACTCTTTATATAGTAAGTATGTAACCAAGAAGAT 92  
 261 ProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThr 280  
 91 CCAACTGTGTCTCTTTACAAGCTATCAAGTCTCGAAGATGATCCCAAC-TCCAAAACA 33  
 281 LysGluPheTyrPalaThrIleLeuAspSerAla 291  
 32 AAGAAATTTGGCGCACCAT-TTGAGATCAGCA 1

RESULT 7  
 BI223892 853 bp mRNA linear EST 11-JUL-2001  
 LOCUS 602941035F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5104113 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI223892  
 VERSION BI223892.1 GI:14677336  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 853)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHAM1251 row: e column: 10  
 High quality sequence stop: 718.  
 Location/Qualifiers  
 1..853  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5104113"  
 /cruise\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /clone\_1lb="NIH\_MGC\_12"  
 /notes="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."  
 BASE COUNT 250 a 175 c 187 g 241 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4e-129 Length: 853  
 Score: 1265.50 Matches: 235  
 Percent Similarity: 96.36% Conservative: 3  
 Best Local Similarity: 95.14% Mismatches: 6  
 Query Match: 50.99% Indels: 1  
 DB: 12 Gaps: 1  
 US-10-070-464-5 (1-465) x BI223892 (1-853)  
 108 GluThrThrAspIleTyrIleAsnIleHisAspIlePheHis-ValPheProGlnSerHis 127  
 61 GAAGAAACAGACATCTGAGTAATAATCAATGACATCTTTTCAGTGTTCCTCCCAAGTCA 60  
 127 sgluglugluileglupheilepheleaserglucylstnrglyphearghisleu 147  
 61 CGAAGGAGAAATGAGTTATTTTGGCTGATGATCAAAAGCTTCGTCATTTATTA 120  
 147 rlysllethrserleleuilegluserytyrlysaraserseglyleuProal 167  
 121 CAAATATCATCTATTTTAAAGAAAGCAATATTAACATCAAGTGGGCTGCTCG 180  
 167 aProserAspPheLysCysProIleLysGluGluIleAlaIleThrSerGlyGluTyrPgi 187  
 181 TCCAAAGTATTTCAAGTGTCTATCAAAAGAGAGATGCAATTTACAGTGGTAATGGGA 240  
 187 uValleuglyArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyrPh 207  
 241 AGTTCTGGCCCGCATGATCTAATATTCAGATTGATGAAGTCAAGAGCTGTATATTT 300

QY	207	EGUICU1yThrlvSAPPSePProLeuGluHhIshIseuYrVla1v1a1SeTyVla1AsnPr	227
Db	301	TGAAGGCACCAAGACTCCCTCTTTAGACATACCTGTACGTACAGTACGTAATCC	360
QY	227	OG1yG1uValThraTgLeuThraSPaTg1yTySeriSseCySeYalSeSerGlnH	247
Db	361	TGGAAGAGTGACAAAGCTGACTGACCGGTGACTACATCTTCTTGATCAGTCAGA	420
QY	247	SCYAspPhePheIleIseTySeriSerAnGlnYsAnP-roHISCYeValSeLeuTy	267
Db	421	CTGTACCTCTTTAATAGTAAGTAATGAAACCAAGAAATCCACACTGTGTCCCTTTA	480
QY	267	TLySLeuSeSerPProG1uAspAspPProthrCYsIyThrlvSG1uPheTPrAlaThrl	287
Db	481	CNAGTATCACACTCCTGGAAGATGACCACTTGCAAAACAAAGAAATTTGGGCCACAT	540
QY	287	eLeuAspSeR1aG1yProLeuPrcAspTyThrProPProG1u1IephSeSPheGluSe	307
Db	541	TGTGATTTACAGACAGTCCCTCTCCGACTATACCTCCACAAATTTCTCTTTGAAAG	600
QY	307	rThrThrlYPhetThrlleuTyTyrGlyMeLeuTyTyrSProHISAspLeuGlnPProGly	327
Db	601	TACTACTGATTTAACTGTAATGTAAAGGATGCTTCAACGCTCATGTACTCAACCTGGAAA	660
QY	327	sLySeTyPProThrValIleuPheIleTyTyrG1yProGlnVal-----Al	342
Db	661	GAAATCTCTACTGTGCTGTCTCATATATGCTGTCTCAGGTGACAGTGTGATATATCG	720
QY	342	a1eAlaG1yAlaProVal	348
Db	721	AGTTAAAGGAGTCCAGTA	739
RESULT 8			
EG259714			
LOCUS	602380072P1	NIH_MGC_92 Homo sapiens	CDNA clone IMAGE:4511080 5',
DEFINITION		mRNA sequence.	
ACCESSION	EG259714		
VERSION	EG259714.1	GI:12769530	
KEYWORDS			
SOURCE			
ORGANISM		Homo sapiens (human)	
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
		1 (bases 1 to 993)	
		NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
		National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished	
		Contact: Robert Strausberg, Ph.D.	
		Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a>	
		Tissue Procurement: ATCC	
		CDNA Library Preparation: Life Technologies, Inc.	
		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)	
		DNA Sequencing by: Incyte Genomics, Inc.	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LNLN at:	
		<a href="http://image.lnl.gov">http://image.lnl.gov</a>	
		Plate: LHAM10393 row: 0 column: 17	
		High quality sequence start: 6	
		High quality sequence stop: 656.	
FEATURES			
Source			
		1..993	
		/organism="Homo sapiens"	
		/mol_type="RNA"	
		/db_xref="taxon:9606"	
		/clone_image="4511080"	
		/tissue_type="embryonal carcinoma, cell line"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_id="NIH_MGC_92"	
		/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;	
		Site_2: SalI; Cloned unidirectionally; clage-dt primed.	
		Average insert size 2.5 Kb. Library enriched for	

full-length clones and constructed by Life Technologies.			
Note: this is a NIH_MGC Library."			
BASE COUNT	296 a	183 c	243 g 271 t
ORIGIN			
Alignment Scores:			
Pred. No.:	5, 41e-122	Length:	993
Score:	1202.00	Matches:	232
Percent Similarity:	90.42%	Conservative:	4
Best Local Similarity:	88.89%	Mismatches:	11
Query Match:	48.43%	Indels:	14
DB:	10	Gaps:	5
US-10-070-464-5 (1-465) x BG259714 (1-993)			
QY	93	11GGLSERVALPROASPSEVALTHRPROLEUIELIETYGUGUUTHRAAPPIE	112
DB	2	ATTGAGTACAGTCCGATTCGTGTGACGCCACCAATTAATTAAGAAACAAACAGACATC	61
QY	113	TRPILEANIIEIASPIIIEPHEIISVALPHEPROGINSERHIAGUGUUIEGU	132
DB	62	TGGATTAATATTCACAGACATCTTTATGTTTTTCCCAAGTCACGAAAGGAATTGAG	121
QY	133	PHIIEPHEIASERGLUCYSLYSTHRGYPHEATGHIISLEUTYLYSIETHSRIIE	152
DB	122	TTTATTTTGGCTTCGAATGCAAAACAGGTTCCCGCATTTATACAAATTAATCATCTATT	181
QY	153	LEULYSGUSELYTYRLYSARGSERGIYGIYLEUPROALAPROSERAPHEIYS	172
DB	182	TTAAAGGAAGCAAAATATTAAGATCCAGTGTGGGCTGCTCCCAAGTGAATTCAG	241
QY	173	CSPROIIIELYSGUGUIIEALIEIETHSERGIYGLUTRPGIUVALEUGIYARGHIS	192
DB	242	TGTCTATCAAAAGAGAGATGACATTACCACTGGATGAGTGGAACTTCTGGCCGGCAT	301
QY	193	GLYSERANIIIEGINVALAPGLUVALYRARGLEUVALTYRPHIIGUICYTHRLYSAP	212
DB	302	GGATCTAATATCCCAATGTGATGAACTCAGAAAGCTGGTATTTTGAAAGCACCAAGAC	361
QY	213	SERPROLEUGIUIEHIISLEUTYRVALIALSERTYRVALIAPROGLYGIUVALTHRARG	232
DB	352	TCCCTTTAGACATCACCTGTACGTAGTCACTTACGTAATCTCGAAGGTGACAAAG	421
QY	233	LEUTHRAAPARGIYTYRSETHISERCYSCYIIIESERGINHISCYASAPHEPHEIE	252
DB	422	CTGACTGACCGGCTACTACATCTTGCTGCTCATGACAGACACTGATCATCTTTATA	481
QY	253	SEILYSTRYSERANGIULYSEAPPROHISCYSVALSEULEUTYLYLEUSERSEPRIO	272
DB	482	AGTAAGTATGATTAACCAAGAAATCCACACTGTGTCTCTTACCAAGTATCAACTGCT	541
QY	273	GIUASAPAPROTHRCYELSETHIRLYSGUHPHETRPALATHRIIEUASPSERIALAGIY	292
DB	542	GAAAGTAGCCCACTTGCAAAACAAAGAAATTTTGGCCACCATTTTGATTCAGACAGT	601
QY	293	PROLEUPROAPRYRTHIRPROPROGIUIIEPESERPHIIGUSERTHIRGYPHETHR	312
DB	602	CCTCTTCTGACTACTCTCCAGAAATTTTCTCTTTGAAAGTACTCTGATTTTACC	661
QY	313	LEU---TYRGIYMETLEUTYLYSEPROHISAPLEUGINPROGLYLYSELYTYRPROTHR	331
DB	662	CTTGCTGGGGGAGT-----CTCTTAACAAAG-----	688
QY	332	VALLEUPHEIETRYR---GIYGLYPROGINVALAIAIEVALAGIYALAPROVALTHRLEU	350
DB	689	GCTTTAATGATCTACAGAGCTGGGAGAGAAATATTTCTTAATGGGCTG---GTCAATPAT	745
QY	351	TRP 351	
DB	746	TGG 748	
RESULT 9	0	*	
BG974587			



Tissue Procurement: T. Rajendra Kumar and Martin M. Matzuk  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 plate: NDCM08 row: b column: 19  
 high quality sequence stop: 551.  
 Location/Qualifiers

## FEATURES

source

```

1.774
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30248202"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_166"
/notes="Organ: brain; Vector: pDR-LIB; Site 1: SfiI
(ggcatatggcc); Site 2: SfiI (ggcgcctggcc);
Non-normalized full-length enriched library 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGCGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 2.05
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Corp. "
```

BASE COUNT 228 a 187 c 160 g 198 t 1 others

ORIGIN

## Alignment Scores:

Pred. No.: 6,41e-114 Length: 774  
 Score: 1127.50 Matches: 221  
 Percent Similarity: 89.88% Conservative: 10  
 Best Local Similarity: 85.99% Mismatches: 19  
 Query Match: 45.43% Indels: 7  
 DB: 14 Gaps: 2

US-10-070-464-5 (1-465) x CB233750 (1-774)

```

51 ThProGluGlyLysTyraLattSerIleLeuLeuAspArgSerGlnThrArgLeuGln 70
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
6 ACTCCAGAGGAGAAACATCGCTGCTCATCTACTACACCGTTCCCGACTGACCTGCGAG 65
71 IleValLeuIleSerProGluLeuPheIleProValGluAspAspValMetGluArgGln 90
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
66 ATAGTTCTGATCTCCCTGAGTTATTCATCCAGTAGAAGATGTCGACAGACAG 125
91 ArgLeuIleGluSerValProAspSerValThrProLeuIleIleTyrgluGlnThr 110
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
126 AGACTTATAGACGTCAGTCTGACTGTGACACCACTGATCATCTATGAAGAAACACA 135
111 AspIleTrpIleAsnIleHisAspIlePheHisValPheProGlnSerHisGluGlu 130
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
186 GACATCTGGATTAATATCCACGATATTTTCATGTTTTCTCAAACTCATGAAGATGA 245
131 IleGluPheIlePheHisAspGluCysValThrGlyPheArgHisLeuTyrylIleThr 150
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
246 ATTAGAGTTATTTTTCCTGTAATGCAAAACAGTTTTCATCTGTTATTAATACACA 305
151 SerIleLeuIleGluSerTyrylLysArgSerSerGlyGlyLeuProAlaProSerAsp 170
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
306 TCATTTTAAAGAGAGCAATATTAACGATCCAGTGTGAGTACTGCCCCCAAGGAT 365
171 PheIleCysProIleIleGluGluIleAlaIleHisSerGlyGluTrpGluValLeuGly 190
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
366 TTCAAGTGTCCTTCAAGAAAGAAATTAACAATTACCGTGTAAATGGAGAACTTGCG 425
191 ArgHisGlySerAsnIleGlnValAspGluValArgArgLeuValTyryPheGluGlyThr 210
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
426 CGGCAATGATCTAATATCTGGGTTGATGAAGCAAGAAAGCTGATCTTTGAAGGACCC 485
```

```

Qy 211 LysAspSerProLeuGluHisHisLeuTyryValValSerTyryValAsnProGluVal 230
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 486 AAGAGCTTCCTTTGGAGGATCATCTGTAGTACCACTATTCAAACCTGGAGAAAGTG 545
Qy 231 ThrArgLeuThrAspArgGlyTyrySerHisSerCysIleSerGlnHisCysAspPhe 250
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 546 GTGAGGCTGACTGACCGGTGCTACTCACACTCTGCTGCTGACCGGCAATGTGATTC 605
Qy 251 PheIleSerTyrySerSerLeuGlnIleLysAsnProHisCysValSerLeuTyryLeu 270
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 606 TTCTAATAGTAGTAGACCAACCAAGAAATCCACTGTGTCTTCCCTTACCAACTCT 665
Qy 270 erSerProGluAspAspProThrCys---LysThrLysGluPheTrpAlaThrIleLeu 288
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 666 CAGTCCCTGAGGATGATCCAGTTTATTAATAAAGAAATTTGGCCACCAATTT 725
Qy 289 -----AspSerAlaClyProLeu-ProAspTyryThrProPro 300
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 726 TGGATTCACCAAGCAAGGTCCTCTTCTGTAAGTACCAACCCCT 768
```

## RESULT 11

BM557438

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM557438 1042 bp mRNA linear EST 20-FEB-2002  
 AGENCOURT\_6578992 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5466759  
 5', mRNA sequence.  
 BM557438  
 BM557438.1 GI:18799430  
 EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1042)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph. D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: DCTD/DRP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 plate: LLCM1967 row: k column: 16  
 high quality sequence stop: 697.  
 Location/Qualifiers

## FEATURES

source

```

1.1042
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5466759"
/cisue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/notes="Organ: skin; Vector: pOT87; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGACAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

BASE COUNT 279 a 223 c 234 g 303 t 3 others

ORIGIN

## Alignment Scores:

Pred. No.: 1.31e-112 Length: 1042  
 Score: 1117.50 Matches: 225  
 Percent Similarity: 65.52% Conservative: 3  
 Best Local Similarity: 64.66% Mismatches: 14  
 Query Match: 45.02% Indels: 106



Db	183	CATCACCTGTAAGTACTAGTTACTTAATCTCTGACAGGTGACAAAGGCTACTGACCGT	242
Oy	237	GLYTYRSerHisSerCysGlyLeuSerGlnHisCysAspPhePheIleSerIleTySer	256
Db	243	GGTACTACACATCTTCTGCTGATCATGTCAGCAGCTGTGATCTTCTTATGAATGATAGT	302
Oy	257	AsnGlnIleAspAspProHisCysValSerLeuTyIleIleuAspSerIleGlyProIleuAspPro	276
Db	303	AACCAAGAAATCCACACTGTGTGTCCCTTTACAAAGCTATCAAGTCTCGAAGATGACCA	362
Oy	277	ThrCysIleThrIleGluPheTyrPheIleThrIleuAspSerIleGlyProIleuAspPro	296
Db	363	ACTTGCAAAACAAAGAAATTTTGGGCCACACATTTGGATTGACAGAGTCTCTTCTGAC	422
Oy	297	TyrThrProProGlnIlePheSerPheGluSerThrThrGlyPheThrLeuTyIleGlyMet	316
Db	423	TATATCTCTCCAGAAATTTCTCTTTGAAAGTACTACGTGATTTACATTTGATGGAGATG	482
Oy	317	LeuTyIleProHisAspLeuGlnProGlyIleTyIleTyProThrValLeuPheIleTy	336
Db	483	CTCTACAAAGCCCTACATCTACAGCCTGGAAAGAAATATCTACTGTCTGTCTATAT	542
Oy	337	GlyGlyProGlnValAlaIleAla-----	344
Db	543	GGTGTCTCTCAGGTGACAGTGTGTAATATCGGTTAAACAGGACTCAATATTTCCGCT	602
Oy	344	-----	344
Db	603	TGAATACCTAGCCCTCTAGTTATGTGGTTGTAGTATAGACAAACAGGGATCTCTGC	662
Oy	344	-----	344
Db	663	ACCGAGGCTTAATTTGAAGCGCCTTAAATATCAATGGGTCAATGAAATTGACGAT	722
Oy	344	-----	344
Db	723	CACGTGAGAGACTCAATTTATCTAGCTTCTCGATATGATATTCAATTGACTAGATCGTG	782
Oy	344	-----	344
Db	783	TGGGCATCACGGGTGTCTTATGAGGAGTACTCTCTCGAATGACATTAAATGACAG	842
Oy	345	-----GlyAlaProValThrLeuTrpIlePhe-Ty	354
Db	843	GTCAGATACCTTCCAGGGGTGCATAAGCTGGGGGCCACAGTCACCTGTGATCTTCTTA	902
Oy	354	AspThrGlyTyrThrGluArgTyrMetGlyHisProAspGlnAsnGlnGlnIleTyTy	374
Db	903	TGATACAGATACACGAAACGTTATATGGGTACCTGAAACAGA---TGAACGGGTATAC	959
Oy	374	IleuGlySerValAlaMetGlnAlaGluIlePhePro-----Se	387
Db	960	TTTCAGGATCTGTG-----CTTGACGCAAAATTTCCCTTGACAATCGTTAATGCCTCAAT	1013
Oy	387	GluProAsnArgLeuLeuLeuHis	396
Db	1014	GGTTCGGGCAAAATCTTGTGCTATCAC	1041
RESULT 13			
LOCUS	CD352521	762 bp	mRNA linear EST 29-MAY-2003
DEFINITION	UI-M-G10-GRK-e-23-0-UI_r1 NIH BMP4_G10 Mus musculus cDNA clone		
ACCESSION	CD352521		
VERSION	CD352521.1	GI:31144032	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 762)		
	NIH-MGC http://mgi.mci.nih.gov/.		

TITLE		JOURNAL		COMMENT	
National Institutes of Health, Mammalian Gene Collection (MGC)		Unpublished		Contact: Robert Strausberg, Ph.D. Email: cga@f5000.nhl.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/mouse1.html This clone was contributed by the Brain Molecular Anatomy Project fBMAP f	
FEATURES		SOURCE		Seq primer: pyx-5.	
location/Qualifiers					
1..762		/organism="Mus musculus"			
		/mol_type="mRNA"			
		/strain="C57BL/6"			
		/db_xref="taxon:10090"			
		/clone="IMAGE: 6856416"			
		/tissue_type="whole brain"			
		/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"			
		/lab_host="DH10B (T1 phage resistant)"			
		/clone_1lb="NH_BMAP_G10"			
		/note="Organ: Brain; Vector: pyx-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Henin Chin, Ph.D., program coordinator."			
BASE COUNT		228 a 158 c 169 g 205 t 2 others			
ORIGIN					
Alignment Scores:					
Pred. No.:		8,59e-107		Length: 762	
Score:		1063.00		Matches: 205	
Percent Similarity:		95.09%		Conservative: 8	
Best Local Similarity:		91.52%		Mismatches: 11	
Query Match:		42.83%		Indels: 1	
DB:		14		Gaps: 0	
US-10-070-463-5 (1-765) x CD352521 (1-762)					
QY	1	ThrGlyThrAlaAsnProLySerValThrPheLysMetSerGluIleMetIleAspAlaGlu	20		
DB	82	ACAGGACAGCAAAACCAAGGCACTTTCAAGAGTGGAGATTGTTGAGCGTCA	141		
QY	21	GlyArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGlu	40		
DB	142	GAGGGGATTATATATCTCATATAGATTAAGAACTGGTTCAACTTCGAGATTCTTTGAG	201		
QY	41	GlyValGlyTyrIleIleAlaArgIleGlyTyrThrProGluGlyLysValTyrAlaTyrSerIle	60		
DB	202	GGAATTGATATTTTCCGAGGCGGATGAGCTCCAGAGGAAACATGCTGCTGATC	261		
QY	61	LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIle	80		
DB	262	CTACTAGACCGTTCCAGACTCACTCGAGATGATTGATCTCCCGAGTTATTCATC	321		
QY	81	ProValGluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerVal	100		
DB	322	CGAGTGAAGATGATCCATGACACAGACAGAACTTATAGAGTCACTTCGACTCTGTG	381		
QY	101	ThrProLeuIleIleTyrGluGluThrThrAspIleTyrIleAsnIleHisAspIlePhe	120		



Db 382 ACACCACTGATCATCTATGAAGAAACACACATCTGGATTAATATCAGATATTTT 441  
 Qy 121 HisValPheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysIys 140  
 Db 442 CATCTTTTCTCTCAAACTCATGAAGATCAATGATTGATTTTTCCTCTGAATGCAAA 501  
 Qy 141 ThrGlyPheArgHisLeuTyrLeuTyrIleThrSerIleLeuYsgIuSerLeuTyrIysArg 160  
 Db 502 ACAGCTTTTCTCTCATCTGTATTAATATCATCTCTTTTAAAGAGACCAATATTAACGG 561  
 Qy 161 SerSerGlyGlyLeuProAlaProSerAspPheIysCysProIleIleGluGluIleAla 180  
 Db 562 TCCAGTGTGGAGTACCTGCCCCCAAGTATTTCAAGTGTCTCTATCAAAAGAAATATAACA 621  
 Qy 181 IleThrSerGlyGluTyrPglIuValLeuGluIleArgHisGlySerAsnIleGluIleValAspGlu 200  
 Db 622 ATTACCGATGGTGAATGGAGAGTACTGGCCGCGCATGTGATCTAA-ATCTGGTTGATGATA 680  
 Qy 201 ValArgArgLeuValTyrPheGluGlyThrIysAspSerProLeuGluIleHisLeuTyr 220  
 Db 681 GCCGAAAGCTGGTGTACTTTGAAAGCACCAAGACCTCTCTTGGACATCACCCTGTAC 740  
 Qy 221 ValValSerTyr 224  
 Db 741 GTGACCAAGTAT 752

RESULT 14  
 AL040398 753 bp mRNA linear EST 29-FEB-2000  
 LOCUS DKEZp434A0714\_r1.434 (synonym: hce3) Homo sapiens cDNA clone  
 DEFINITION DKEZp434A0714\_5', mRNA sequence.  
 ACCESSION AL040398  
 VERSION AL040398.1 GI:5409350  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 753)  
 AUTHORS Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
 TITLE EST (Koehler, et al.)  
 JOURNAL Unpublished  
 COMMENT Contact: Koehler K  
 MIBS  
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by BMPZ (Biomedical Research Center at the Charite,  
 Berlin/Germany) within the cDNA sequencing consortium of the German  
 Genome Project.  
 No 5' sequence available.  
 This clone (DKEZp434A0714) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 FEATURES  
 source location/Qualifiers  
 1..753  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKEZp434A0714"  
 /issue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="434 (synonym: hce3)"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 218 a 150 c 170 g 214 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,47e-106 Length: 753  
 Score: 1055.00 Matches: 193

Percent Similarity: 99.49%  
 Best Local Similarity: 98.97%  
 Query Match: 98.97%  
 DB: 9 42.51%  
 US-10-070-464-5 (1-465) x AL040398 (1-753)

Qy 149 IleThrSerIleLeuYsgIuSerLeuTyrIysArgSerGlyIleuProAlaPro 168  
 Db 1 ATTACATCTATTTTAAAGAAAGCAAAATATTAAGATCACTGATGGCTCTGCTCTCA 60  
 Qy 169 SerAspPheIysCysProIleIleGluGluIleAlaIleThrSerGlyIleuProAlaPro 188  
 Db 61 AGTATTTTCAAGTGTCTCTATCAAAAGAGAGATACCAATTACCAAGTGTGAATGGAGATT 120  
 Qy 189 LeuGlyArgHisGlySerAsnIleGluIleValAspGluValArgArgLeuValTyrPheGlu 208  
 Db 121 CTTCGCGCGCATGTGATCTAATATCAAGATTGATGATGACAGAGCTGTGATATTTCGA 180  
 Qy 209 GlyThrIysAspSerProLeuGluIleHisLeuTyrValValSerTyrValAsnProGly 228  
 Db 181 GGACCAAGAGCTCCCTTTAGAGCATCACCTGTACGTACGTACGTAAATCTCGCA 240  
 Qy 229 GluValThrArgLeuThrAspArgGlyTyrSerHisSerCysIleSerGluHisCys 248  
 Db 241 GAGGTGACCAAGCTGACTGACCTGCTGCTACTACATTTTCTGCTCATCAGTACGACTGT 300  
 Qy 249 AspPhePheIleSerIleTyrSerAsnGluIleAsnProHisCysValSerLeuTyrIys 268  
 Db 301 GACTCTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 360  
 Qy 269 LeuSerSerProGluAspAspProThrCysIleThrIysGluPheTyrPheIleu 288  
 Db 361 CTATCAAGCTCTGTAAGTGAAGCCCAACTTGCAAAAGAAATTTTGGCCACCATTTTG 420  
 Qy 289 AspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSerThr 308  
 Db 421 GATTACAGAGTCTCTCTCTGACTATACCTCTCAAAATTTCTCTTGAAGTACT 480  
 Qy 309 ThrGlyPheThrLeuTyrGlyMetLeuTyrIlePheProHisAspLeuGluProGlyIleIys 328  
 Db 481 ACTGATTTTACATTTGATGGAGAGCTCTACAGAGCTCATATCTACAGAGCTCTGAGAGAGAA 540  
 Qy 329 TyrProThrValLeuPheIleTyrGlyGlyProGluValAlaIle 343  
 Db 541 TATCTTACTGTGCTGTTCATATATGTGTCTCTCAGTCAAGTTG 585  
 RESULT 15  
 BG974324 831 bp mRNA linear EST 12-JUN-2001  
 LOCUS BG974324  
 DEFINITION BG974324.1 GI:14361961  
 ACCESSION BG974324  
 VERSION BG974324.1 GI:14361961  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 831)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch  
 Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov



Plate: LLM10978 row: e column: 20  
High quality sequence stop: 797.  
Location/Qualifiers  
source  
1. .831  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4979707"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_11b="NCL CGAP Mam4"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."

BASE COUNT 235 a 191 c 190 g 215 t  
ORIGIN

Alignment Scores:  
Pred. NO.: 3.46e-104 Length: 831  
Score: 1040.00 Matches: 190  
Percent Similarity: 93.75% Conservative: 5  
Best Local Similarity: 91.35% Mismatches: 13  
Query Match: 41.90% Indels: 0  
DB: 12 Gaps: 0

US-10-070-464-5 (1-465) x BG974324 (1-831)

```
QY 136 AlasergluCyslysthrGlypheArgHisleuTyrlslethrSerlleuLysglu 155
    |||||
Db 6 GCGTCCGGGAGGCGGTGGGTTTCGTCATCTGTATAAATCACAATCATTTTAAAGGAG 65
QY 156 SerlystYrlylaArgSerSetGlylyleuProalaProSeraSpPheLysCysProile 175
    |||||
Db 66 AGCAAAATATAAACCGTCCAGTGTGAGTACTCTGCCCAAGTGTTCMAAGTGTCTTATC 125
QY 176 LysglugluIleAlaIleThrSerGlygluTrrpGluValleuGlyArgHisGlySeraSn 195
    |||||
Db 126 AAAGAGAAATTAACAATTACCAAGTGTGATGGAGAGTACTTGGCCGCAATGATCTAT 185
QY 196 lIeGlnValaspGluValArgArgleuValTyrphegluglyThrLysaspSerProleu 215
    |||||
Db 186 ATCTGGGTTGATGAAGCCAGAAAGCTGTGTACTTGAAGGACCAAAAGACTCTCCTTG 245
QY 216 GluHisHisleuTyrlValSerTyrlValasnProGlygluValThrArgleuThrAsp 235
    |||||
Db 246 GAGCATCACTGTGTACGTACCAAGTATGCAAACTCTGAGAAAGTGTGAGGCTGACTGAC 305
QY 236 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerlystYr 255
    |||||
Db 306 CGTGGCTACTACACTCTCTGCTGCTCAGCCGCACTTGTGACTTCTTCAATAGTAAGTAC 365
QY 256 SerAsnglnLysAsnProHisCysValSerleuTyrlLysleuSerSerProGluAspAsp 275
    |||||
Db 366 AGCAACGAGAAATTCACACTGTGTGTCCTCTACAAACTCTCAAGTCTCGAGGATGAC 425
QY 276 ProThrCyslysthrLysgluPheTrrpAlaThrIleleuaspSeraIaglyProleuPro 295
    |||||
Db 426 CCAAGTTATATAAACAAAGAAATTTGGGCCACCATGTGAGATTCAGCAGGTCCTCTTCT 485
QY 296 AspTyrThrProProGluIlePheSerPhegluSerThrThrGlyPheThrLeuTyrgly 315
    |||||
Db 486 GACTACACCCCTCCAGAAATTTTCTTTTGAAGTACTACTGATTTACACTGTATGGA 545
QY 316 MetleuTyrlLysProHisAspLeuGlnProGlyLyslystYrProThrValleuPheIle 335
    |||||
Db 546 ATGTTGTAATAAGCCTCATGACCTACAACTGGAAGAAATACCCCACTGTGTATTCTATA 605
QY 336 TyrglyglyProGlnValAlaIle 343
```

Db 606 TATGAGGTCCCAAGTGCACACTG 629  
Search completed: October 16, 2003, 03:25:35  
Job time : 2818.69 secs